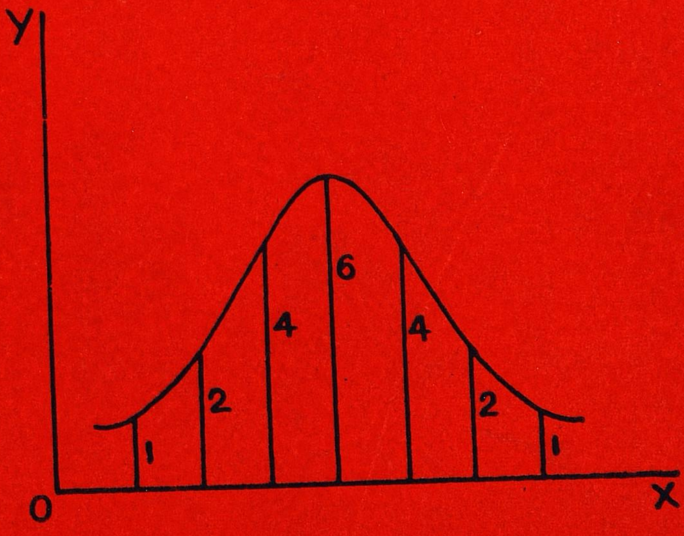
STATISTICS

HIGHER SECONDARY FIRST YEAR

VOL. I



TAMILNADU TEXTBOOK SOCIETY

STATISTICS

Vol. I

Higher Secondary – First Year

TAMILNADU TEXTBOOK SOCIETY
MADRAS

First Edition—1980

Editorial Board Chairman
(Author & Review Committee Members)

Thiru. **M. Sankaranarayanan,** M.A., B.Sc.,
Joint Director of Statistics,
Department of Statistics,
MADRAS - 600 006.

REVIEW COMMITTEE MEMBERS:

Thiru. **T. K. Manickavachagam Pillai,** M.A., L.T.,
Professor of Mathematics (Retd.)
A.C. College of Technology;
MADRAS - 600 035.

Thiru. **R. Hanumantha Rao,** M.A.,
Professor of Mathematics,
P.S.G. Arts College,
COIMBATORE.

Price: Rs. 8-00

This book has been printed on concessional paper of 60 G.S.M. substance made available by the Government of India.

Printed at
SANKAR PRINTERS, MADRAS - 600 018.

CONTENTS

FIRST YEAR : : FIRST PAPER

CHAPTER I

INTRODUCTION

In recent years, the science 'Statistics' has advanced to a very great extent. In fact, it is being widely used in all branches like Astronomy, Mathematics, Economics, Physics, Chemistry, Biology, Genetics, Agriculture, Veterinary, Engineering, Medical, Business etc. and more particularly in every research department. Therefore, it has become necessary and also it is profitable to study statistics.

Meaning of the word 'Statistics'

The term 'Statistics' is used in two senses. In the narrow, common sense, the word 'Statistics' denotes some numerical data. In other words, it can refer to facts which can be expressed in terms of numbers as in the case of 'unemployment statistics' or 'industrial accidents' in a country.

In the wider sense, the term 'Statistics' refers to the statistical principles and methods, which have been developed for handling numerical data. We are going to study this aspect in this book.

Origin of the word 'Statistics'

The origin of the word 'Statistics' may be traced to the Latin word 'Statius' which means State. This shows that statistics has had some relationship to the affairs of the State. Any Government requires facts about the manpower available especially for the defence of the country, facts about the financial and economic conditions of the people for taxation, facts about the number of people and facts about the production of food grains etc. Hence, statistics are described as by-products of the Government administration.

Definition and characteristics of Statistics

The task of formulating even a preliminary definition of the term 'Statistics' is complicated because the conception or nature of 'Statistics' differs widely. Many authors have defined this term in different ways.

Numerical statement of facts

Statistics has been defined as numerical statements of facts in any department. But mere figures are not necessarily statistics. For example, the numbers 25, 30, 40 and 50, though they are quantitative figures, do not mean statistics since they do not mean anything else. However, if these figures represent the age of persons like 25 years, 30 years, 40 years and 50 years or weight of persons like 25 kg., 30 kg., 40 kg., and 50 kg., they will become statistics. In such situations, these numbers show some relationship with the people concerned. Hence, it is argued that **Statistics must be capable of being placed in relation to each other and they should be comparable.**

Statistics are aggregates of facts

It may be noticed that statistics relate to relationship to the people. Fact or information about one person will not constitute statistics, since it will represent a mere statement as the age of a particular person is 20 years or the weight of a particular person is 50 kilograms. Such a single figure cannot be compared. Therefore, facts about a large body of persons or individuals will alone represent statistics. Hence, statistics are defined as aggregates of facts.

Statistics are quantitative

Statistics must be capable of being expressed in quantitative terms. Even quantitative expressions like male or female, young or old, worker or non-worker, married, widowed or unmarried etc. will become statistics only if they are expressed in terms of figures as follows:

The number of male population in a village	=	750
The number of female population in a village	=	745
The number of workers in a village	=	400
The number of non-workers in a village	=	725
The number of married persons in a village	=	350
The number of widowed persons in a village	=	25
The number of unmarried persons in a village	=	150

Statistics are affected to a marked extent by a multiplicity of causes

Generally, facts and figures are affected to a considerable extent by a number of forces operating together. The statistics of prices are affected by conditions of supply, demand, imports, exports, taxation, circulation of money and a large number of other causes. It may also be difficult to study the effect of each of these causes separately on the general price level.

Ensuring reasonable standards of accuracy in Statistics

Facts and figures about any phenomenon are obtained in two ways, either by actual counting or by estimation. But estimation cannot be as accurate as the actual countings. However, a reasonable standard of accuracy is necessary in the estimates also, in order to draw valid conclusions from them.

Statistics as Department of Enquiry

Another interpretation of the term statistics is that it relates to a Department of Enquiry. In this field, the statistics indicate the relationship between the individuals in respect of the characteristics under study and hence the statistics should be comparable.

Statistics as Science of counting

Prof. A.L. Bowley has given different definitions of statistics. One of his definitions is that statistics is a science of counting. This is too narrow a definition as it relates to only one aspect of statistics namely collection of data. Since the collection of data is only a means to an end and not an end in itself, this definition

is very limited. Besides collection of data, it also involves analysing, interpreting and drawing inferences from them. Hence this definition is not complete and therefore not satisfactory.

Statistics as Science of Averages

Another definition of statistics given by Prof. A.L. Bowley is that it is a Science of Averages. The data collected from an enquiry have to be condensed and summarised into certain measures or parameters. The most important of these measures is average or mean since he defines statistics as science of averages. But there are other measures like measures of dispersion or measures of correlation which are also vitally important as Average. Therefore, this definition of statistics, as the science of averages, is also not complete and satisfactory.

Statistics as a Science of measurements of social organism

Prof. A.L. Bowley also defines 'Statistics' as a science of measurements of social organism' regarded as a whole in all its manifestations. This definition limits the scope of statistics tc man and his activities. But modern statistics deal not only with social phenomena but also biological, physical phenomena and hence this definition is also not satisfactory.

Statistics as a science of estimates and probabilities

Boddington has defined statistics as the science of estimates and probabilities. As estimates and probabilities are only part of statistical methods, this definition is also not complete.

Croxton, Cowden and Seligman defined statistics as collection, presentation, analysis and interpretation of numerical data to throw light on any sphere of enquiry.

Thus, a definition of statistics, comprehensive enough, to accommodate the above views is given by Secrist as follows:

> "Statistics are aggregates of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable

> standards of accuracy; collected in a systematic manner for a pre-determined purpose and placed in relation to each other".

Statistics are not confined to one branch of study. A systematic study involving methods of collection, analysis and interpretation of quantitative data affected by a number of causes in any field of enquiry can be termed as statistics.

Main Divisions of Statistics

Any modern science is generally divided into two kinds namely 'Pure Science' and 'Applied Science'. Similarly, statistics as a Science has been divided into two divisions namely, Statistical Methods and Applied Statistics.

Statistical Methods

Statistical Methods are concerned with the formulation of general rules and principles applicable in different branches like methods of collection of data, classification, tabulation, analysis, comparision by means of averages, diagrams, co-efficients and correlation and finally interpretation of data.

Applied Statistics

Applied Statistics deals with the application of these statistical rules and principles to concrete subject matters like wages, prices, trade, population etc. Applied Statistics may consist of 1. Biometry, (dealing with measurement of living organism), 2. Psychometry (which deals with mental problem), 3. Vital Statistics (dealing with birth and death of human population) 4. Administrative, social and economic statistics.

Applied Statistics can be further divided into two kinds: (1) Descriptive Applied Statistics and (2) Scientific Applied Statistics.

Descriptive Applied Statistics

It deals with data which are already available relating either to the past or to the present. For example, business statistics

which deal with analysis, measurement and presentation of business data relating to the past or present can be classified as Descriptive Applied Statistics.

Scientific Applied Statistics

This deals with the formulation of scientific laws on the basis of quantitative data collected for descriptive purposes by means of suitable statistical methods. Business forecast on the basis of data collected in the past can come under this category.

Statistics as Science and as Art

The recent advances in statistical theory make us consider statistics as a science of inductive reasoning from observed numerical facts. But there are others who argue that statistics is not a science but it is only a scientific method.

Science deals with the building of knowledge and Art deals with the application of that knowledge. Hence the statistical methods are considered as a part of science and the Applied Statistics is treated as Art. Thus, statistics is considered both as a science and an art.

Functions of Statistics

The chief and ultimate function of statistics is to increase our knowledge. It performs six main functions as detailed below:

1. Condensation.
2. Comparison.
3. Correlation.
4. Correction.
5. Evaluation of results.
6. Enlargement of experience.

1. Condensation

Statistics simplifies a mass of unwieldy and complex data. The capacity of the human mind is limited. If a large number of data are presented, it will be always difficult for the human mind to know their significance. But complicated data can be condensed into totals, averages and they can be presented by means of graphs and diagrams, which can be easily understood by human mind without much strain. Statistics are of great help in these methods.

2. Comparison

The essence of statistics is not only counting but also comparison. Data on certain items by themselves do not serve any purpose. They must be capable of being compared with similar data of other places or period of time. The price of a particular commodity in a place does not serve any purpose, unless it is capable of being compared with the price of the same commodity in some other place or in some other period.

3. Correlation

Another important function of statistics is to find out the relationship between two or more phenomena, as in the case of supply and demand or application of fertiliser and agricultural production. For these purposes, the statistical measures generally used are co-efficient of correlation, co-efficient of variance and co-efficient of association etc.

4. Corrects vague notions

Sometimes our notion about certain items may be either vague or indefinite due to want of correct data. Instead of merely saying that the cost of living has increased, we can precisely say, with the help of cost of living index number, that the cost of living has risen by certain definite measurable and understandable quantity.

5. Evaluation of the effects

Statistics also help to evaluate the effect of implementation of certain programmes or policies in the country. For example,

the effects due to the introduction of prohibition on the socio-economic conditions of the people can be evaluated by means of statistics.

6. Enlargement of Experience

Statistics makes it easier to understand, explain and measure the effects of some actions on others, as in the case of reaction of the rise in the price of gold on the prices of other commodities.

IMPORTANCE OF STATISTICS

At present, there is an increasing influence of statistical methods in every field of activity. The reasons for such influence are many. Modern civilisation requires more specific information than what was necessary in the past and this is achieved only through statistical methods. Precise information are required for forecasting the weather both for the purpose of war and peace. Nowadays, even the trend of public opinion is being measured by statistical methods.

Another reason for the growing importance of statistics is the increase in the spread of modern scientific spirit. People do not accept any phenomena merely on faith or trust. They try to explain complex phenomena by determination exactly of their component factors. Therefore, quantitative data are inevitably required in seeking exact scientific knowledge.

In certain fields of Enquiry, the laboratary methods of science are not applicable. Conclusions have to be drawn from careful analysis of data of practical experiences. In such cases inferences have to be drawn from large masses of quantitative data. Thus, statistical analysis has come to play an essential part in the use of scientific methods for the extension of knowledge.

Importance to State

Modern Governments require numerical data on a variety of subjects like production, prices, economy, births, deaths,

diseases etc. for taking effective steps and introducing welfare measures.

Importance in Economics

Statistical data and methods are useful in understanding the economic problems and also help to formulate economic policies. Statistics are more useful in different branches of Economics, namely consumption, production, value and distribution.

Importance in Planning

Economic Planning of a country without a variety of statistical information is an imperfect and impossible task. After implementing the programmes, statistical methods are quite necessary to evaluate the success of the plan.

Importance in business

Statistics is an aid to commerce and industry. It helps the businessmen and industrialists to decide things more precisely and objectively. In the face of keen competition, the businessman, before launching his plan, should have detailed statements regarding the demand for his goods, production by other manufacturers. Success in a business depends on accurate forecasts of sales and statistical methods are very helpful in these forecasts.

Cost Accounting is entirely a statistical approach and with the help of this technique, the business people can decide the price of his products. In order to have a steady demand for the goods manufactured, the quality of the goods should be ensured. In this attempt, the Statistical Quality Control plays an important part in standardising the goods produced. Operational Research is another branch of statistical methods, which helps the businessmen to reduce the cost of production and maximise the production.

Utility in Insurance Companies

Statistics is extensively used in insurance companies. Statistics helps them to find the expectation of the life of the people and thereby fix the premium to be paid by the policy holders.

This is very essential from the business point of view, in view of the keen competition.

Uses in Banking Institutions

Statistical methods are useful to bankers and brokers in Stock Exchanges. It helps bankers to have knowledge about the seasonal variations in the credit demands. Similarly, the brokers too require information on the fluctuations in the investment and market trend to have a good business.

Researches

Research is an important part in any activity, since without research no improvement can be effected. Improvement is essential especially in a dynamic world. Statistics is an indispensable tool in the field of research in any branch of study. Statistical methods and techniques are widely used in these researches.

Statistics and Social studies

Any Government, which wants to introduce welfare measures, should know the magnitude of the social problem to be tackled and statistics help them to measure this.

Other fields

Statistical methods are applicable to all sciences like Astronomy, Physics, Chemistry, Biology, Meteorology, Geology, Medicine, Engineering etc.

Limitations of Statistics

At present, statistical methods are extensively used in all fields of scientific investigation. Hence, great care has to be taken in handling and using statistics.

It does not reflect individuals

Statistics deals with aggregates of facts. Hence, it will not reflect the real position of an individual.

Statistical rules are true in the long run

Statistical Laws and Rules are true only in the long run.

Applicable to quantitative problems

All statistics are numerical statements of figures and hence, they can be adopted to situations capable of being expressed in quantitative terms. Qualitative characteristics like poverty, richness etc., can be studied only when they are expressed in quantitative terms.

Statistics is not the only method of studying the problem

Statistical methods cannot always provide solution for all problems under study. It is one of the various methods.

Need for homogenity in the data

Effective study of the problem can be ensured only when the data represent a homogeneous population. In other cases, the results may mislead.

Statistics are estimates rather than actuals

Statistical calculations are based more on estimates-rather approximates, than actuals.

Statistics without context may mislead

If statistics are quoted without context, it may mislead to give false conclusions.

Liable to mislead

Statistics can be misused without knowing the purpose for which and the circumstances under which it was collected. Hence, it can confuse instead of convince. Therefore, it is a dangerous tool in the hands of inexperienced persons.

Distrust in Statistics

In spite of the great usefulness of statistics, distrust in statistics also prevails. It has been humurously classified along with lies. This kind of distrust prevails not only among illiterates but also among the highly educated people. The reasons for such situation are many.

There are certain people who have some strong and firm views about certain policies. If the results on the basis of statistical surveys do not support their views, they will condemn statistics. On the other hand, if they find that these data support their pet theories, they will praise statistics. They are nothing but opportunists.

There are another category of people who do not understand the techniques of statistical methods. If there is a slight difference in the values arrived at by two independent enquiries conducted by two persons they would not hesitate to condemn it. They do not know that though these two people have selected samples from the sample universe, the samples selected by each person themselves may differ and consequently there would be a difference in the result. Because of this, statistics is otherwise known as science of variation. People, who are not aware of this practical situation, may condemn it.

Sometimes, people used to quote statistics out of context in support of their arguement. For example, a politician may quote the Cost of Living Index Numbers of two Countries and say that the cost of living in one place is higher than that in other place. This is not correct. The real picture is that the cost of living index mumbers gives a comparison of the same place to its past.

Inspite of these distrusts, limitations, statistics has come to play a vital part in all our activities. Refinements are being made in the statistical methods due to researches and advanced studies.

Exercise

1. Define 'Statistics'.
2. Write an essay on the term 'Statistics'.
3. 'Statistics is a science of Averages' - Discuss.
4. 'Statistics is a science of Counting' - Discuss.
5. Discuss the limitations of Statistics.
6. Discuss the merits and demerits of Statistics.

CHAPTER II

STATISTICAL DATA

Collection of data

Collection of data is the first step in any statistical investigation. Statistical data are of two kinds namely (i) primary data and (ii) secondary data.

Primary data

The data, which are collected for the first time, are called primary data and they are original in character.

Secondary data

Secondary data are those, which have been collected, tabulated and presented in some form by some one else. Secondary data are second-hand data. The data, that have been published by Government or by research agencies, are primary data for the agencies, who have collected them and they serve as secondary data for those, who use them for their studies. For instance, the data collected by Government during the population census are primary data to the Government, whereas they are secondary for those who are using them afterwards. The distinction between primary and secondary data is, therefore, one of degree only.

Primary data have certain advantages over the secondary data. They are truthful and purposive. The disadvantages are that their collection is time-taking and expensive. Therefore, it is always better for an individual investigator, before undertaking a study, to find out whether any data are already available on the subject under study. But care has to be taken in handling the primary data as to the concepts and definitions adopted in

the collection of primary data. A well thought-out planning for the actual collection of data will reduce the errors to a minimum and thus saves time, money and labour.

Methods of collecting primary data

Statistical data are usually obtained in three ways, namely Census method, Registration method and Mailing the questionnaire.

In the Census Method, the desired information from the individuals are collected by interviewing each informant. Decennial population census, Quinquennial livestock census conducted in this country come under this category. In Registration Method, as for instance in the recording of births, deaths and marriages, the Registration authority will generally be a Government official empowered for the purpose. The method is more effective and satisfactory, since failure to register birth and death will be a penal offence. In the case of Mailing questionnaire, required data are collected by sending a questionnaire and instructions to individuals with a request to furnish the details called for in the questionnaire. This is sometimes known as Correspondence method. Collection of statistics relating to consumption of raw materials and production etc. from industrial units will come under this category.

Census Method

Each method of collecting the statistical information has its own merits and demerits besides certain limitations. Hence each method can be adopted only in certain specific fields of circumstances.

The census method of compiling data can be adopted by agencies, which can command the necessary finance and get the required number of enumerators. This method is generally adopted by Government and similar influential bodies. The first step in such a process is planning the whole scheme sufficiently early. The information required are obtained as the answers to a series of questions, which the enumerators would put to the informant. These questions are contained in a form known as schedule or questionnaire.

Schedules and Questions

Having selected the points or items, for which answers are required, each question is formulated or arranged in such a way as to be readily and accurately answered.

Questions

The questions in a schedule or questionnaire can be classified into four types.

1. **Specific information questions:** These questions require only specific answers as in the case of age of a person or number of members in a family.

2. **Open type questions:** These questions do not require any specific answers. On the other hand, the informants can give reply as he desires. For example, one can give his opinions about family planning as one feels.

3. **Alternate type questions:** The answers to this type of questions are either this or that. For example, the sex of a person may be either Male or Female. Sometimes the answers to certain questions may be either 'Yes' or 'No.'

4. **Multiple choice questions:** This type of questions will have multiple choice of answers. For example, answers regarding the marital status of a person may be any one of the following:

Single; Married; Widowed; Divorced; Separated.

Leading questions, which prompt informants to answer questions in a particular manner, should be avoided.

In most cases, unskilled persons are chosen as enumerators and so, these enumerators should be trained and instructed in the process of obtaining answers to the questions in the schedule. Generally, each enumerator will be supplied with an instruction book.

Schedules

The preparation of the schedules has to be done with great care and caution. The following points should generally be adhered to:

(1.) Each question in the schedule should be simple and as clear as possible. (2) Questions, which are unduly inquisitive and which are likely to give offence or which are likely to irritate the informant, should be avoided. (3) Answers should be subjective and capable of being tabulated. (4) Instructions and definitions should be concise. The enumerator and the informant should never be in doubt as to what exact information is required to a particular question. (5) The units used in the statistical study should be clearly stated, (6) The questions should be arranged in a logical order and in such a manner as to allow sufficient space for the people to answer. (7) Leading questions should be avoided, (8) Before finalising the schedule, it is desirable to pre-test it on a group of persons and revise it if any short comings are discovered. (8) The enumerator should be specially instructed to explain the informant the purpose of the statistical enquiry and solicit not only information but also his co-operation and help. (9) It is also better to plan the interview in such a way that the interview takes minimum time and yields maximum results. A model questionnaire for enquiry of the income of families is given below.

Questionnaire for enquiry of the income of families.

I. General

1. State
2. District
3. Village
4. Door No.
5. Name of the head of household
6. Extent of agricultural holding : Wet
   (Hectares) Dry

II. Composition of the family

S.No.	Name	Sex	Age	Relationship to the head of family	Occupation
(1)	(2)	(3)	(4)	(5)	(6)

III. Income of the family

Source	Annual Income (Rs.)
1. Agriculture	
2. Industries	
3. Receipts from outside	
4. Others (specify)	
Total	

Registration Method

As explained earlier, in this method the informants are expected by an Act of the Government to record voluntarily the occurence of the event and generally a Government servant will be empowered for this purpose. This system is generally more effective, since failure to record the data will be considered as a penal offence. Of course, ignorance of the law and the necessity of recording on the part of the informant will be a source of error in this method. But such errors will occur only in the initial stage and in course of time the scope and the extent of this type of error will diminish, as experience and knowledge grow. One great limitation in this method is that all types of data cannot be collected. This method is generally resorted to for collecting information regarding events taking place occassionally. At present Vital Statistics such as birth and death are being recorded in this manner. It is also the least expensive.

Correspondence Method

Sending a schedule or a questionnaire by mail is less expensive method of collecting data. The informant will have more time at his disposal to answer the various questions in the schedule. Sometimes, confidential information, which a person would hesitate to divulge to an enumerator may be given in writing, if the informant is sure that his identity will remain unknown.

On the other hand, a large proportion of persons would fail to make any reply even an addressed or stamped cover is enclosed. Under such circumstances, it is logical to expect that the answers received would be more from persons of particular type such as those who are educated, and who have understood the significance and use of the Enquiry. Therefore, the results obtained may not represent the whole population, from which information are sought. Another defect is that often incorrect answers are given either due to ignorance or due to wilful negligence. In any case it is preferable to enclose with the schedule a letter of explanation seeking co-operation and help from the informants. When only a part of the schedule is finally received, it is necessary to make sure that no selective factor is present.

Nowadays, information are being collected mainly through correspondence method. This is particularly so in the field of Industries. In order to overcome the defect due to non-response, special steps are being taken to make the non-response as a penal offence. Hence, this system can be more effectively employed by Government authority with the help of necessary legislations.

Secondary Data

So far, we studied about the collection of primary data from the individual members or individual units in the population. In certain cases, secondary data, either published or unpublished, are also available. In such cases, compilation of data, rather than collection, is involved. Though there is some advantage, by way of saving in time, energy, cost etc. in the use of secondary data, great care has to be exercised in the use of secondary data.

Sources of secondary data

The sources of secondary data may be divided into two kinds, namely published and unpublished. Even the published data can be classified into three kinds such as (1) those published by Government in their official publications, (2) those published by semi-official or quasi-Government institutions, and (3) those published by private agencies.

Official publications

These include the publications of State Governments, Central Government, International Bodies like United Nations Organisation, Food and Agricultural Organisation, World Health Organisation, International Monetary Fund etc.

Semi-Official publications

The publication of various corporate bodies, like Municipalities, Corporations, Quasi-Government undertakings like Agricultural Finance Corporation, Life Insurance Corporation, Reserve Bank of India etc. come under this category.

Private publications

The publications of various Chambers of Commerce, Mill-owners' Associations, Research Institutes may come under this category.

Precaution in the use of secondary data

It is not always safe to take the secondary data at their face value without cross examination. It is always advantageous to check the secondary data in the light of the following points before adopting them.

1. The agency, which has collected the data and its object in collecting those data.
2. The sources, from which the data are collected.
3. The methods adopted for the collection of data.
4. The concepts and definitions adopted for the various terms in the schedules.

5. Type of supervision exercised at the time of collection to ensure accuracy.
6. Mode of tabulation either mechanical or manual.
7. Comparability of the data collected with previous sets of figures, if any.
8. Period of collection, and the period to which the data relate and the period of publication of results.
9. The size of the population covered.

It is always preferable to have primary data rather than secondary data because of the above limitations.

STATISTICAL DATA AND THEIR KINDS

Statistical data are of two kinds, qualitative and quantitative. In the case of qualitative characteristics, the individuals or members or units under consideration are distinguished by some quality or attribute. But in the case of quantitative characteristics, the individuals are distinguished by measurements or counts.

Qualitative

Information on (1) Colours of flowers such as red or yellow, (2) Sex of persons such as Male or Female, (3) Marital status such as married, widowed, separated, never married etc. (4) Economic status of persons such as earning independent, earning dependent, dependent etc., (5) Type of buildings such as tiled or thatched etc., are some of the examples coming under qualitative items. Qualitative information are also known as Attributes.

Quantitative

The number of members in a family, length of a road, income of individuals, weight of the yield, height of persons, duration of employment etc., can be classified as quantitative. But details either qualitative or quantitative can be expressed in terms of numbers. The following may illustrate the position:

Colours of flowers:

Number of red flowers
Number of white flowers and
Number of yellow flowers in the garden.

Sex:

Number of male members or
Number of female members in the family.

Marital Status:

Number of married persons
Number of divorced persons and
Number of widowed persons in the society.

Economic Status:

Number of earning dependents
Number of earning independents and
Number of dependents in the family.

A characteristic such as income, weight, height, number of members in a family, the production of a factory, capable of being expressed in terms of quantitative values, can be classified into two kinds such as Continuous and Discontinuous or Discrete.

Continuous values

Quantities which can take any numerical value within any range are called continuous values. The weight of a person can be either 56 kg. or 56.51 kg. etc. In all such cases, what we require is only sensitive instrument to record the details in terms of minute units and there is a possibility of having persons with a weight equal to 56.51 kg.

Discontinuous or Discrete values

Quantities, which can take only discrete values, are called discontinuous or discrete values.

The number of workers in a factory, the number of boys in a class room, the number of students who have passed the examination etc. are one and the same type of information. That is, information, which can be expressed in terms of whole numbers such as 0, 1, 2, 3 etc. and on no account they would take a value between 2 and 3 or 3 and 4, or 7 and 8. In other words, they cannot have a value as 2·35, 2·80, 3·15, 4·50. In such cases, there will not be any continuity between two consecutive full numbers. Hence, these types of information are known as **discrete values.**

Variable

A measurable characteristic, the value of which can vary from one individual to another, is called a variable.

Let us consider the characteristic, namely, the weight of persons. The weight of persons will vary from individual to individual and it cannot be a fixed value for all the persons. Therefore, weight, a characteristic which can be expressed in quantitative value and the value of which varies from person to person is called a variable. The height of persons, wages of workers, yield of land etc. are only variables. A characteristic, which is variable in nature from person to person, or from place to place or from time to time is called a variable.

Constant

A value, which is a fixed quantity for all the members, is called a constant. Example—The number of head for a person is for all individuals.

Representation of constants and variables

In Statistics, generally, the variables are represented by the letters z, y, x, w, v, u, ie. letters read from the end to the beginning. The constants, which take only a definite fixed value irrespective of the individuals, are represented by the letters read from the beginning to end such as a, b, c, d etc.

Exercise

1. Explain the various stages in a statistical investigation.
2. What are the preliminary steps to be taken before the collection of data?
3. What is a questionnaire?
4. Prepare a model questionnaire for conducting an enquiry of the income of families.
5. Define primary and secondary data.
6. What are the different methods for the collection of primary data?
7. Define secondary data. What are the precautions to be taken in using the secondary data?

CHAPTER III

CLASSIFICATION OF DATA

The statistical data are not collected just for the sake of data, instead, they are collected for detailed study and to arrive at certain conclusion about the present and formulate future policies. We cannot make out anything from a mass of data if they are presented as they are. If we want to have any meaningful interpretation, the mass of data should be classified according to our needs. Therefore, classification of statistical data is the most important part in any detailed analysis. The process of arranging data in groups or classes according to resemblance or similarities is technically called classification. Classification can be done according to our needs and convenience. Nevertheless, certain important points have to be followed in the classification. We can study this in detail when we learn the construction of frequency tables.

STATISTICAL TABLES

Tabulation

Tabulation is one of the methods of presentation of statistical data. Tabulation is a scientific process involving presentation of data in an orderly manner, so as to bring out its essential features and chief characteristics. The main object of tabulation is to present voluminous and heterogeneous statistical data in a condensed and homogeneous form.

A statistical table is a systematic organisation of statistical data in columns and rows. It can be either qualitative or quantitative.

Qualitative classification

When classification is done on the basis of a qualitative variables it is called Descriptive or Qualitative classification. We

can classify the total number of persons in our area according to the sex, namely males and females. We can also classify them according to the literacy, ie, literates and illiterates. Such a classification based on one characteristic is called a simple or two-way classification.

Sex	No. of persons
Males	125
Females	120
Total	245

Literacy	No. of persons
Literates	95
Illiterates	150
Total	245

Sometimes, a classification can be done on the basis of two or more characteristics simultaneously. Such a classification is called manifold classification. For example, the number of people in a village can be tabulated on the basis of their sex and literacy as follows:

Sex	Literates	Illiterates	Total
Males	60	65	125
Females	35	85	120
Total	95	150	245

Quantitative classification

Classification of data according to some quantitative variables is said to be quantitative classification. The number of families in a town can be classified according to the size of the family, that is, the number of members in a family as follows:

Size of the family	No. of families
2 and less than 2 members	25
3 members	45
4 members	70
5 members	15
above 5 members	5
Total	160

When the data are grouped into classes of appropriate size showing the number in each class, we get what is known as frequency distribution. The frequency distibution replaces the larger number of original measurement by a fewer number of frequencies.

Generally, the tabulation can be either simple or complex based upon the number of characteristics involved in the classification. A simple table gives information about one characteristic whereas, the complex table gives information about more than one characteristic.

Statistical tables can be classified as one way tables or two way tables or three-way tables or manifold tables depending upon the number of characteristics involved simultaneously in preparing the table.

A table may be primary table or derivative table. A primary table is the one prepared on the basis of the actual data collected as follows:

Details of Expenditure	Actual expenditure
	Rs.
1. Food	150
2. Clothing	50
3. Education	15
4. Fuel and Light	25
5. Miscellaneous	10
Total	250

Instead of giving the actual expenditure, the expenditure on different items can be given as a percentage of the total expenditure as follows:

Details of expenditure	Percentage of Expenditure
1. Food	60
2. Clothing	20
3. Education	6
4. Fuel and Light	10
5. Miscellaneous	4
	100

A table prepared on the basis of derivatives like percentage, ratio, and proportion is called derivative table.

Construction of tables

Most of the statistical tables present quantitative information in a classified manner, in rows and in columns. A statistical table has four important parts.

1. Table number and the title.

2. Stub (the headings of the rows or the heading portion of the rows).

3. Caption or box-head (the column heading portions or the heading portions of the columns).

4. Body.

It may contain foot note and a note indicating the sources of the data.

Title and Identification Number

Every table should have a number at the top, (such as table 3 or 4 or 4·1) for easy identification. It should also have a title. (Numbers of the table should be given in a serial order.) The wording of the title should be brief and at the same time explanatory.

Arrangements of items in Stub and Caption

Depending upon the importance of the statistical data, they may be arranged either geographically, or chronologically or alphabetically or in magnitude.

Sources

It is always customary to give the source of the data so that any one, who uses them, may refer to the original data, if needed.

A specimen of the table prepared is given below:

Table No. 1.1

Trend in population growth in 1901 - 1971

Year	Tamil Nadu		India	
	Population (millions)	Index	Population (millions)	Index
(1)	(2)	(3)	(4)	(5)
1901	19.25	100.0	238.34	100.0
1911	20.90	108.6	252.01	105.7
1921	21.63	112.4	251.24	105.4
1931	23.47	121.9	278.87	117.0
1941	26.27	136.5	318.54	133.7
1951	30.12	156.5	360.95	151.4
1961	33.69	175.0	439.07	184.2
1971	41.10	213.5	547.37	229.7

(Source: Computed from data available from 1961 and 1971 Census).

Rules to be observed in the construction of tables

1. The tables should be simple, precise and easy to understand.

2. It is always advantageous to have more than one table, if the data are very large. In such cases single table would confuse the users.

3. The title of the table and the headings and sub-headings of rows and columns must be simple and self-explanatory.

4. The number of main headings should be few as far as possible.

5. The number of rows should also be few.

6. The units of measurements should always be given.

7. Figures may be approximated or rounded off before tabulation so as to avoid strain in reading and understanding.

8. If the figures are very large, they may be rounded either in thousands or in lakhs.

9. Sufficient space may be given between each row.

10. Columns should be numbered for easy reference.

11. Items in rows can be arranged either alphabetically or chronologically or geographically according to size or magnitude of the data depending upon the importance.

12. The total at the end of the columns and rows should be given.

However, no hard and fast rules can be given. Depending upon the need, occassion and nature of data, tables can be constructed.

Advantages of tabulation

1. Tabulation enhances clarity and gives answers to the questions at a glance of the table.
2. Tabulation helps the analysis and interpretation of the data collected.
3. Errors can be easily noticed if the data are tabulated.
4. As the data in a table are arranged in a logical order, it helps the reader to have a logical reference of the data.
5. It facilitates easy comparison.

In order to prepare a table, classification of the data collected is very important.

Classiffcation according to class-intervals

One of the important classifications that we come across in statistics is the classification according to class intervals. Numerical characteristics are classified by giving arbitrary limits. The age or height or weight of persons vary from person to person. The entire range of ages or heights or weights can be split up by arbitrary boundary lines and those units which are almost alike in respect of a particular character can be put together in one class.

If the ages of a particular group of persons vary from 22 to 47 years and if it is decided to divide this group into 6 smaller groups, we can have the boundary lines fixed as 25, 30, 35, 40, 45, and 50 years. These points of boundary lines are called *class limits*. The group constituted by two consecutive class limits is known as a class. The interval or range or distance between two consecutive limits in known as *class interval*. The number of observations coming under a particular group is known as frequency of that class.

In the above example, we can group together all the persons with ages 20 years and more but less than 25 years in the class 20-25. Similarly, all the persons with age 25 years and above but less than 30 under the group 25-30. Similarly, we can have the remaining groups as 30-35, 35-40, 40-45, and 45-50. In this example the magnitude of the class interval is 5 years.

As far as possible the magnitude of the class intervals should be uniform so that all statistical measures can be computed without much difficulty. Generally, the class intervals may be either 5 or multiples of 5 such as 10, 25, 50, 100, 250, 500 and 1000 and so on depending upon the magnitude of the data to be tabulated. The number of groups or classes to be formed should not be too small or too large. Generally it is preferable to have the total number of classes varying from 8 to 15 and in very exceptional cases involving large number of data we can have even 20 classes.

There may not be any, frequency in a particular class or group. Even on such condition, the particular class or group should be retained so as to avoid errors in computation.

Exclusive method

In the above example, the classes 20-25 and 25-30 years are expressed according to exclusive method. The upper limit os a particular class is the same as the lower limit of the succeeding class. In this case, one unit with a value 24.99 years will come under the class 20-25 and another unit with exactly 25 years under the class 25—30 years. However, the difficulty in the interpretation of the class limits can be avoided by having the classes as follows: 20 years and under 25 years, 25 years and under 30 years and so on.

Inclusive method

Class intervals can also be expressed by the inclusive method. In this case, the upper limit of a class will not be the same as the lower limit of the succeeding class. In the above example, we can have the class as 20-24, 25-29, and 30-34 and so on. In this case, the first class namely 20-24 includes all items between 19.5 to 24.5 years. But the inclusive method is not generally adopted because of the break in continuity, since the portion 29-30 is omitted.

Exercise

1. What is meant by classification of data?

   Explain the objects and characteristics of classification.

2. What is the procedure for classifying observations? What points are to be considered while tabulating them?

   Describe the different tables in use.

3. What are the chief functions of tabulation? What precautions would you take in tabulating the data?

4. Define a frequency distribution.

5. The following figures are the rainfall in millimetres recorded for 60 days in a station. Classify the data with appropriate class intervals.

50, 53, 24, 75, 40, 28, 75, 51, 13, 39, 99, 80, 16
14, 48, 43, 11, 56, 85, 79, 64, 28, 10, 95, 63, 28
12, 69, 75, 64, 15, 96, 48, 63, 29, 14, 87, 69, 33
20, 88, 61, 20, 48, 19, 23, 24, 98, 65, 34, 90, 70,
45, 60, 22, 75, 55, 14, 17, 18.

CHAPTER IV

DIAGRAMMATIC OR GRAPHIC REPRESENTATION

We have so far seen that statistical data are short forms of presenting the state of affairs; and classifications of the data in the form of tables are still shorter forms of presenting the figures. Now we can see that graphs and charts are the shortest methods of presenting data given in the form of tables. One great advantage of this is that even at a glance of the charts, without going deep into the details of the figures, under various columns, one can have a rough idea about the nature of the distribution. Even a layman, who is not well versed with facts and figures, can understand the matter. Comparison of different distributions can also be made easy. In short, graphic representations are the quick, easy and effective means of attracting the attention.

But, there are certain basic differences between the tables and charts or graphs. In the case of charts, many sets of facts as given in the different columns or rows of the tables cannot be given. Though exact figures are given in the tables, the figures in the charts are generally approximated by rounding them off to the nearest tens or hundreds or thousands, or lakhs etc. Lastly, preparation of charts requires certain amount of labour and time and yet their effctiveness is indeed a great advantage.

Types of Charts

Charts are of different kinds. However, they may be divided into curves or line diagrams, Bar charts or Bar diagrams, Pie-diagrams indicating area diagrams, Pictographs etc.

Plotting a Curve

The curve or line diagram is formed by joining different points. The different points representing statistical data are plotted with reference to two intersecting lines called the axes. The horizontal line is generally known as the X-axis and the vertical line is called Y-axis. The point of intersection of the two axes is called 'origin' generally indicated by '0'.

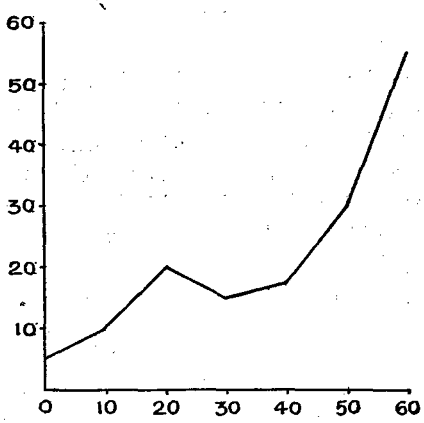


FIG. 1

Line Graph

Whenever chronological data are depicted by means of curve, the periods are shown on the X-axis and the values of the variable are shown on the Y-axis. In other cases, the classes which represent the value of the variables are shown on the X-axis and the frequencies corresponding to the classes are shown on the Y-axis. But an important point to be noted is that instead of the classes, the mid-values of the classes are also marked on the X-axis on the assumption that the frequencies within a particular class is uniformly distributed or in other words, they are centered around the mid-value or the central value of the classes, as if they are having the same value.

Frequency curves (or) Curves for frequency distribution

Generally, Histograms, Frequency Polygons, Frequency curves or Pie-diagrams are used for frequency distributions. These diagrams are generally known as area diagrams.

Histograms

Rectangles with area proportional to the frequencies of the classes are erected over each of the classes. Since the widths of the classes or the class intervals of all the classes are the same, the area of the rectangles erected over each class will be proportional to the height of the rectangle. Hence, the height of the rectangle over each class will be proportional to the frequency of the respective class. The area of all the rectangles will represent the sum total of the frequencies of all the classes.

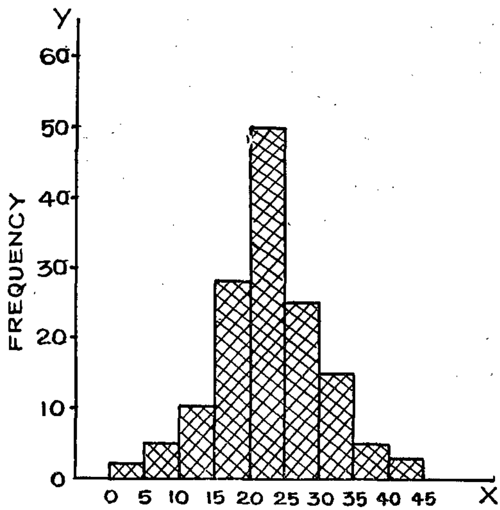


FIG. 2
Histogram

Frequency Polygon

The histogram can be converted into a polygon by joining the mid-points of the tops of all the rectangles by means of straight

lines. The assumption underlined is that the values in a class are uniformly distributed and consequently, the mid-value represents the class. Since this polygon having many sides, generally more than 5 sides, represents the frequencies, it is called frequency polygon.

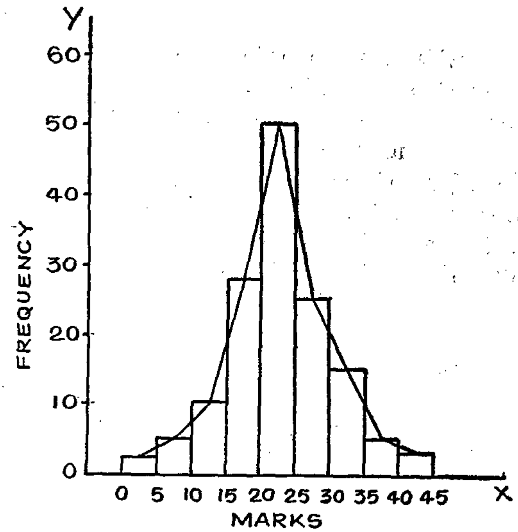


FIG. 3
Frequency Polygon

Frequency curve

If the mid-points of the tops of all the rectangles are joined by means of a smooth curve, instead of by straight lines as in the case of frequency polygon, we will get a smooth curve called frequency curve. If the class interval is reduced smaller and smaller, each of the rectangles erected over each class willl be approximately a straight line and the polygon obtained by joining the tops of these straight lines over all the classes will be a smooth curve rather than a polygon. The ordinate at the mid-point of each interval is proportional to the frequency of the variable in that interval. The area between two ordinates is proportional to the frequency between the corresponding values of the variables.

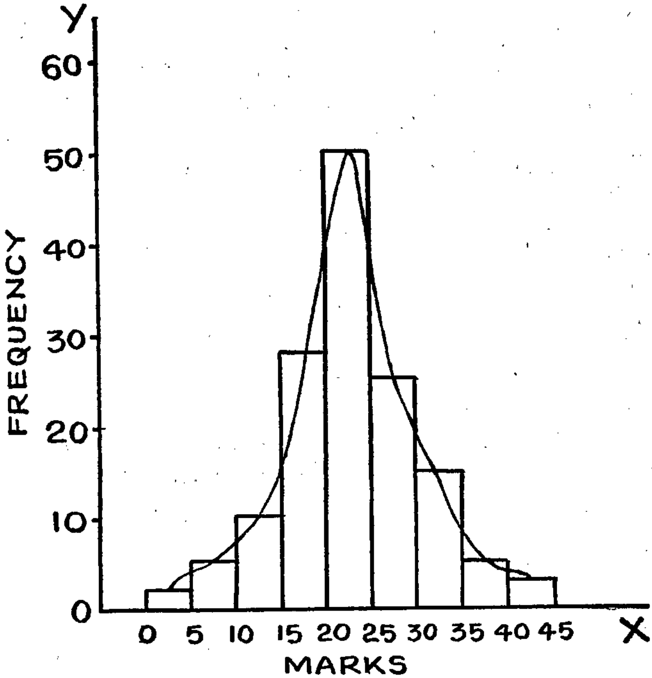


Fig. 4.
Frequency Curve

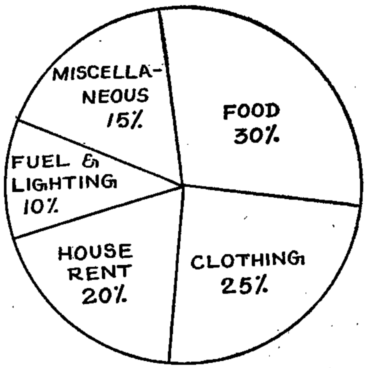


Fig. 5
Pie Diagram

Pie diagram

Pie diagrams will be in the form of a circle. A circle will consist of 360° and hence the total of the frequencies in all cases will be represented by 360°. Therefore, the total of 360° will be divided among the different groups proportional to the respective frequencies of the groups or classes (Fig. 5).

Bar diagram

There are different types of bar diagrams such as simple bar diagram, Multiple bar diagram and component bar diagrams.

1. Simple Bar Diagram

This is the simplest of all the diagrams. It is used to represent one variable at a time. Production of commodity can be represented by means of simple diagram (Fig. 6).

Commodity		Production (in tonnes)
Rice		5302,000
Cholam		464,000
Cumbu		273,000
Ragi		315,000
Other commodities		333,000
	Total	6687,000

2. Multiple diagram

When two or more values are to be compared, it is better to construct two or more bars side by side depending upon the number of values under comparison. If males and females are to be represented, they can be represented by means of bars drawn side by side (Fig. 7).

District	Males	Females	Total
		(lakhs)	
Madras	12.97	11.72	24.69
Chengalpattu	14.93	14.15	29.08
North Arcot	19.05	18.50	37.55

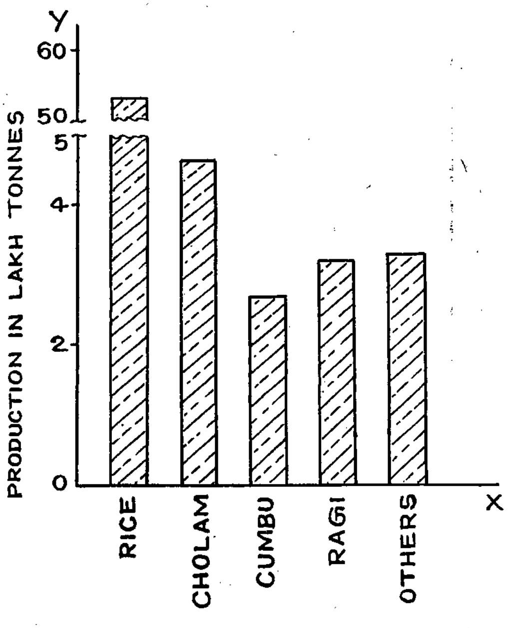


FIG. 6
Bar Diagram

3. Component Bars

These bar diagrams can be used to represent the various components in a particular item. For example, the total population may consist of males and females. One bar containing different portions each representing each of the constituents may be drawn (Fig. 8.).

District	Males	Females	Total
	(in lakhs)		
Madras	12.97	11.72	24.69
Chengalpattu	14.93	14.15	29.08
North Arcot	19.05	18.50	37.55

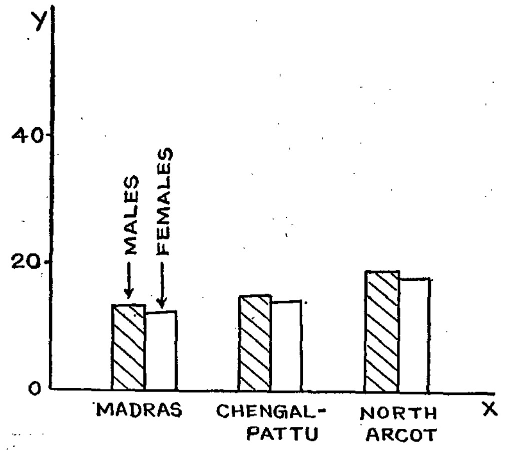


FIG. 7

Multiple Bars

In certain respects, the bar diagrams resemble the Histograms. But there is one basic difference between these two. In Histogram, the rectangles constructed will be a continuous series and there will not be any space between two adjacent rectangles, whereas in the case of bar diagrams there will be space between the two bars to have a clear distinction between them. Generally, bar diagrams are adopted whenever there is no frequency distribution. When the chart is expected to convey a general impression, simple bar charts without the use of a scale on the X-axis may be drawn. But equal gaps should be provided between one bar and the other

bar. Bars may be either horizontal or vertical. However, scale is also very useful. The length of the bar should be proportional to the magnitude of the data.

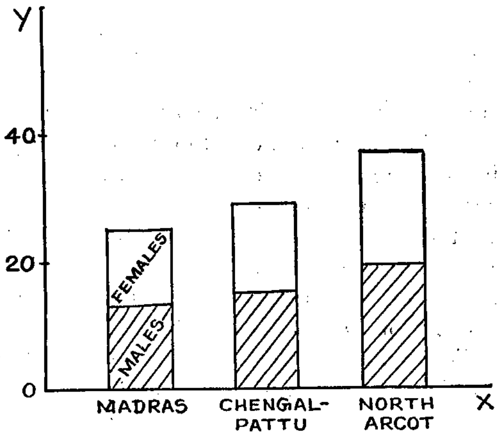


FIG. 8
Component Bars

Pictographs

In this, pictures are drawn to represent the details. For example, the production of cycles or cars can be shown by means

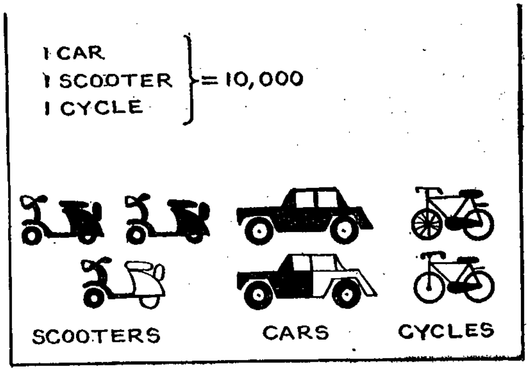


FIG. 9
Pictographs

of the pictures of cycles or cars, respectively. Here too, the quantity can be represented by the number of pictures to be drawn and a scale can also be adopted namely 1 picture represents either 1000 or a lakh of cycles or cars as the case may be (Fig. 9.).

Ogive

This is only a line curve. This curve is used for representing the cumulative frequencies. A smooth curve approximating the cumulative polygon is the cumulative frequency curve known as 'Ogive'. There are two curves of this type, one for representing 'Less than cumulative frequency' and the other for 'more than cumulative frequency'. These two curves may intersect at a point. A perpendicular from the point of intersection of these two Ogive curves can be drawn to the X-axis. The point of intersection of the perpendicular and the X-axis will give the Mean or Median value of the variables which will be explained in the next chapter.

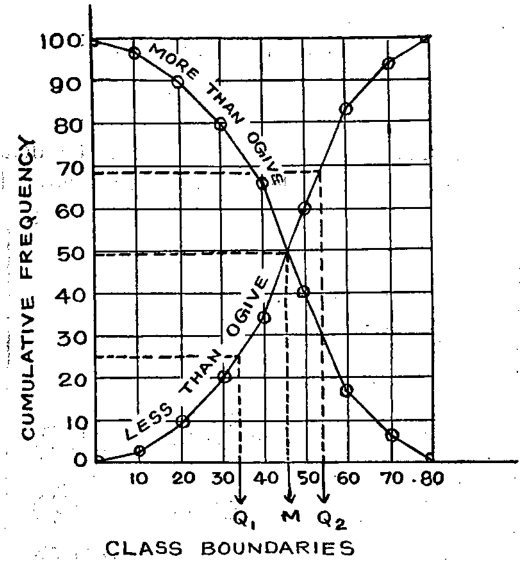


FIG. 10

Ogive Curve

Lorenz curve

Lorenz curve is the name given to a variant of the percentage cumulative distribution when both variates are expressed as percentages. It serves to measure evenness of distribution of two variates i.e. to find for instance if values less than 10 percent of one variate correspond to values less than 10 percent of the

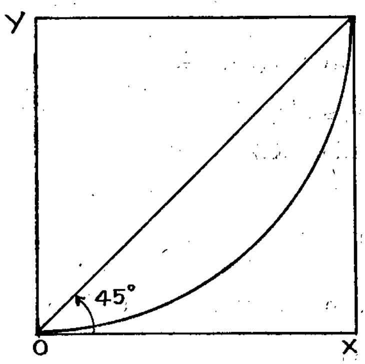


FIG-11

Lorenz Curve

other variate. It is often employed to measure concentration of wealth or income. Even distribution of wealth will correspond to the diagonal of the figure given above. Any departure will be measured by the concavity or the convexity of the curve.

Exercise

1. What is the need for diagrammatic representation of data?

2. Write short notes on any three important methods used for diagrammatic representation.

3. Discuss the various methods of diagrammatic representation of data.

4. How are frequency polygons and frequency curves constructed?

5. Describe an ogive curve. How is the median located?

6. Write short notes on Lorenz Curve.

7. The following table shows the monthly expenditure of three families. Represent the data by means of suitable diagram on percentage basis.

		Family 1	Family 2	Family 3
i.	Food articles (Rs)	45	85	53
ii.	Clothing (Rs)	20	30	17
iii.	Education (Rs)	20	12	15
iv.	Rent (Rs)	20	25	19
v.	Others (Rs)	10	18	16

8. Draw a pie diagram for the following data.

Crop	Area in hectares
Paddy	2564
Cholam	817
Cumbu	450
Ragi	320
Other cereals	446
Pulses	403

CHAPTER V

FREQUENCY DISTRIBUTION

In any statistical enquiry, the data collected will be arranged, analysed and interpreted. The data collected may be used as such without any editing or arranging. This process may be known as analysis of ungrouped data, since the data collected are not grouped or classified. On the other hand, in most cases, the data collected will be classified or grouped first and they will be analysed. This process is said to be analysis of grouped data.

The data may be grouped in two ways either in discrete manner or in a continuous manner. Generally, the classifications may also be known as frequency distribution. They may be either discrete frequency distributions or continuous frequency distributions.

Let us examine an illustration giving the monthly salary of 20 persons.

275, 240, 275, 250, 225, 240, 275, 240, 225, 250, 225, 260, 250, 240, 225, 250, 240, 225, 220, 250.

These data are called ungrouped data.

1. Discrete frequency distribution

In this method, we have to find out the number of persons getting a particular specified salary and arrange them as follows:

1 person is getting Rs. 220.

5 persons are getting Rs.225 each.

5 persons are getting Rs.240 each.

5 persons are getting Rs. 250 each.

1 person is getting Rs. 260.

3 persons are getting Rs. 275 each.

The above details can be put in the form of a table as given below:

Salary (Rs.)	No. of persons
220	1
225	5
240	5
250	5
260	1
275	3
Total	20

This is called a frequency distribution. Since all the values are only whole numbers, it is also called a Discrete frequency distribution.

Continuous frequency distribution

Instead of counting the number of persons getting a particular specified salary, we can fix a salary range and find the number of persons getting a salary in a particular range. This can be done in the following manner. We can have the salary range as Rs. 200-225, Rs. 226-250 and Rs. 251-275.

Pay range (Rs.)	No. of persons
200—225	6
226—250	10
251—275	4
Total	20

The ranges we have adopted are continuous without any break.

Various steps are involved in grouping the raw data or in grouping the ungrouped data into a frequency distribution either discrete or continuous. However, if the ranges of the distribution are continuous, we can use it for both discrete and continuous distributions. We shall examine the construction of a frequency distribution in detail with the help of an illustration.

FREQUENCY DISTRIBUTION

Raw data or ungrouped data

The statistical data, we generally collect, are raw data or ungrouped data. Let us consider the marks obtained by 50 students in a paper on Statistics (Maximum marks 150).

67,	75,	127,	80,	85,	83,	93,	97,	91,	98,
98,	94,	102,	100,	102,	104,	105,	105,	102,	103,
121,	114,	79,	72,	82,	87,	88,	98,	107,	103,
90,	92,	98,	118,	111,	110,	106,	97,	109,	108,
107,	76,	89,	85,	88,	97,	91,	98,	112,	106.

The above figures are nothing but raw or ungrouped data since these data are recorded as they occur without any preconsideration. These raw data occur in a haphazard manner without any order and hence it is difficult to make any comment on them. Hence, these values have to be arranged in some order, either ascending or descending, so as to help the analysis. These can be arranged in the following ascending order:

67,	72,	75,	76,	79,	80,	82,	83,	85,	85,
87,	88,	88,	89,	90,	91,	91,	92,	93,	94,
97,	97,	97,	98,	98,	98,	98,	98,	100,	102,
102,	102,	103,	103,	104,	105,	105,	106,	106,	107,
107,	108,	109,	110,	111,	112,	114,	118,	121,	127.

Array

An arrangement of items in an ascending order or in a descending order of magnitude is called an array. The array helps us to see at once the maximum and the minimum values. It also gives a rough idea of the distribution of the items. The formation of an array is of a very difficult, tedious and cumbersome job. Especially, this is so when we have a large number of data. Hence, the condensation of the array in a more useful and easy manner is essential.

Frequency

A close scrutiny of the data considered in the above example will show that the same value repeats or occurs more than once. Each of the values 85, 88, 91, 103, 105, 106 and 107 are repeated twice: the values 97, and 102 are repeated thrice and the value 98 is repeated five times. The remaining values are occuring only once. These values are more frequented or they can said to be frequented twice, thrice or five times as the case may be. In other words, the frequency of the repetition of the same value can be said as 2 or 3 or 5. Each of the remaining values occurs only once. Hence the number of times a particular value occurs or repeats is called the frequency of that value.

In the above case, the frequency of the values 85, 88, 91, 103, 105, 106 and 107 is 2; the frequency of 97 and 102 is 3; and the frequency of 98 is 5. The frequency of the remaining values is 1 each. This can be arranged as follows:

Value	No. of times (or frequency)
85	2
88	2
91	2
97	3
98	5
102	3
103	2
105	2
106	2
107	2
For the remaining values each	1

Discrete frequency distribution and its formation

A table formed with two columns, one for the value of the item and the other for its frequency is called a discrete or discontinuous frequency distribution or simply a frequency array.

Formation

The given values can be arranged first in an ascending order. The first value can be written in the column for the value. The occurrence of the value can be indicated by a mark (/) called tally mark against the value in the frequency column. The second value of the array can be considered next and the value can be first entered in the value column just below the one already written and its occurance can be denoted by a tally mark as before in the frequency column against it.

If the second value considered happens to be one and the same as the first value, another tally mark can be made in the frequency column against the first value itself (//). instead of entering the value once again. In this manner all the values can be entered in the value column and their respective frequencies can be indicated by tally marks against them in the frequency column.

The number of tally marks against a particular value will indicate the number of times the particular value occurs or repeats itself. If the number of tally marks against a particular value is large, the counting of the number of tally marks itself will be a tedious problem. This difficulty in counting the tally marks can be reduced by grouping the tally marks into clusters of 5 tally marks each(~~////~~). Every fifth tally mark can be drawn across the previous four tally marks sothat these five tally marks will constitute a cluster of 5 tally marks. In this process we can count the number of clusters and arrive at the correct number of tally marks by multiplying the number of cluster by 5. The total number of tally marks against all the values or the total of all the frequencies given against each value will be equal to the total number of items or total number of values. A frequency table can be constructed as follows for the values given in the example below. (marks of students):

67, 67, 67, 71, 71, 71, 71, 71, 75, 75,
77, 78, 79, 79, 79, 79, 79, 79, 79, 80,
80, 82, 82, 85, 85.

Value	Tally marks	Frequency
67	///	3
71	~~////~~	5
75	//	2
77	/	1
78	/	1
79	~~////~~ //	7
80	//	2
82	//	2
85	//	2
		25

We can have certain details about the students. Only two students have secured the maximum marks of 85 and three students

have secured the minimum marks of 67. The marks secured by the maximum number of students is 79 and their number is 7. These are the advantages of classification. This type of condensation can be adopted only when the values are repeated more than once. If the different values occur only once this type of condensation may not be useful and some other type may be thought of.

A better method is to divide the data into classes or groups and then count the number of students falling under each of the groups as in the case of discrete frequency distribution. This is known as continuous frequency distribution.

Continuous frequency distribution

Let us consider a series of data of marks secured by students in a particular paper in an examination. As the maximum marks for a paper is 100 and the minimum is 1, we can divide this range 0 - 100 into 5 or 10 convenient groups as follows:

1—10, 11—20, 21—30, 31—40, 41—50 and so on upto 91—100.

Let us study about certain characteristics of the frequency distribution.

Class: Each group is called a class.

Class interval: The range or difference between the maximum and minimum value of a class is known as class interval. Generally class interval will be uniform for all the classes.

Class limits or class boundaries: The lower value of a class interval is called the lower limit or lower boundary and the higher value of the class interval is called upper limit or upper boundary.

Mid-value

The central value of the each class is known as mid-value. Generally in most cases, the average value between the lower limit and the upper limit will be the mid-value. The difference between

two successive mid-values will be equal to class interval. There are two methods namely inclusive method and exclusive method.

Inclusive method

In the group 1 - 10 of the above example, we can accommodate not only 1 and 10 but also the values between 0 -10. In this manner we consider the other classes also. The values included in each class interval are as follows:

1. The value equal to the lower limit of the class interval. or
2. The value lying between lower limit and the upper limit of the class interval. or
3. The value equal tc the upper limit of the class interval.

Not true class intervals

The inclusive method class intervals are also called not-true class interval. In this method the upper limit of a particular class is not equal to the lower limit of the succeeding class. There will be a gap between two successive class intervals. We can have this type of class intervals only when we have discrete values. If the value is continuous as in the case of weight cr age it will be very difficult to adopt inclusive method since we cannot adopt intermediate values (say 25·7) between two integral values (say 25 and 26).

True class intervals

In the case of continuous variables, we take the class intervals in such a way that there is no gap between any two class intervals. The class intervals are fixed in such a way that the upper limit of each class is equal to the lower limit of the next class. For example, the class intervals are fixed as 0-10, 10-20, 20-30 and so on. This type of class intervals are known as true class intervals.

One difficulty may arise in this type of class interval. If a particular value is exactly equal to the upper limit of a class

say 20, difficulty may arise whether to include this value in th class 10-20 or 20-30.

In such case the following assumption is made.

If the value is exactly equal to the upper limit of a class, it will be included only in the next higher class and not in that class. The value 20 will be included in the class 20-30 and not in the class 10-20.

Thus in the case of true class intervals the values to be included in a particular class are as follows:

1. Values equal to the lower limit of the class interval.

2. Values greater than the lower limit and less than the upper limit of the class interval.

In the case of true class intervals the class limits are known as true lower limit or true lower boundary and true upper limit or true upper boundary.

Exclusive Method

In the case of these class intervals, values which are equal to upper limit of a class are not included in that class and instead they would be included in the next class. The values to be included in a particular class are decided with reference to the lower boundary of the class interval. The upper boundary is not at all taken into consideration or in other words it is always excluded from the consideration. Hence this method is called exclusive method.

Conversion of Not-true class intervals into True class intervals

Generally, true class intervals are adopted. In case the class intervals are arranged in the inclusive form, they should be converted into exclusive form before analysis is undertaken. The following points may be considered for conversion.

(1) Find out the difference between the upper limit of the first class and the lower limit of the second class. (11—10=1).

(2) The difference has to be divided by 2, which is equal to $\frac{1}{2}=0.5$.

(3) This result (0.5) has to be subtracted from each of the lower limits of the class intervals. Then the resultant lower values of the respective classes will be equal to 0.5, 10.5, 20.5, 30.5, 40.5 etc.

(4) The result obtained by division of the difference by 2 as in process (2) has to be **added** to the upper limit of each of the classes.

(5) The resultant upper limit would be 10.5, 20.5, 30.5, 40.5, 50.5 etc.

The given class intervals and corresponding true class intervals of a particular distribution are given below:

Given class interval	True class interval
1–10	0.5–10.5
11–20	10.5–20.5
21–30	20.5–30.5
31–40	30.5–40.5
41–50	40.5–50.5
51–60	50.5–60.5
61–70	60.5–70.5
71–80	70.5–80.5
81–90	80.5–90.5
91–100	90.5–100.5

Width of class interval

The difference between the true upper limit and true lower limit of the class is called the width of the class or shortly known

as class interval. In this case, the width of the class interval is uniformly same for all classes and it is said to be **equal class interval.**

Mid-value or Class mark

The value, which is mid-value between the lower limit and the upper limit, is called the mid-value or class mark or central value. This is calculated as follows:

The sum of the upper limit and lower limit is divided by 2 to arrive at the class mark.

Class interval	Class mark
0.5 — 10.5	$\frac{0.5 + 10.5}{2} = \frac{11}{2} = 5.5$
10.5 — 20.5	$\frac{10.5 + 20.5}{2} = \frac{31}{2} = 15.5$
20.5 — 30.5	$\frac{20.5 + 30.5}{2} = \frac{51}{2} = 25.5$

In this case the difference between two successive class marks will be equal to the class interval itself.

Thus it may be seen that in the case of equal class intervals the following conditions would be satisfied.

1. The differences between the lower limits of the successive classes is equal to the class interval.
2. The differences between the upper limits of the successive classes will be equal to the class interval.
3. The differences between the mid-values or class marks of two successive classes would be equal to the class interval.

When a set of values of a variable are grouped into several classes and all these classes are arranged in the order of magnitude

with the corresponding frequency against each class, the result is a frequency distribution or frequency table.

Construction of continuous frequency distribution

The construction of a continuous frequency distribution involves two important steps.

1. Choice of class interval and consequently choice of class limits.
2. Recording the frequency of each class.

Let us consider this with an example.

The following data give the net weight of each bag of grain obtained from the harvest of plots of uniform size. The weights are recorded in kilograms.

67,	75,	127,	80,	85,	83,	93,	97,	91,	98,
98,	94,	102,	100,	102,	104,	105,	105,	103,	102,
121,	114,	79,	72,	82,	87,	88,	98,	107,	103,
90,	92,	98,	118,	111,	110,	106,	97,	109,	108,
107,	76,	89,	85,	98,	97,	91,	98,	112,	106.

Choice of class intervals and class limits

There are no hard and fast rules in the choice of class intervals and consequently no hard and fast rules in the number of classes to be provided. However, the number of classes should not be either too large or too small. If it is too large, the very advantages of having frequency tables in condensing the data would be defeated. On the other hand, if the number of classes is too few, the frequencies and consequently the values would be crowded in a class, which would again cause inconvenience. Hence, the number of classes to be provided should depend upon the nature of data. But it would be advantageous, if the number of classes ranges between 8 and 15.

The number of classes to be provided depends upon first **on the Range ie:** the difference between the highest and the lowest

values in the series of data. Therefore, we have to first select the highest value and the lowest value and then find out the difference. Depending upon the number of classes we can fix a class interval by dividing the range by the appropriate number of classes desired and the quotient obtained can be taken as the class interval.

In our example 127 kg. is the maximum value and 67 kg. is the minimum value. The range in this case is 60. If we want 10 classes, we can have 6 kg. (60/10) as the class interval. On the other hand, if we want to have 12 classes we can have 5 (60/12) as the class interval. But it is always advisable and also convenient in computation, if the class interval is taken either as 5 or multiples of 5 such as 10, 20, 25, 50, 100, 200, 250, 500, 1000 as the case may be.

In this example, we shall have 10 as the class interval. Since the minimum value is 67, we shall have the first class as 61-70 and the other classes as 71-80, 81-90, etc. and the last class will be 121-130. In this case, the real classes would be as follows:

60·5 — 70·5

70·5 — 80·5

80·5 — 90·5

90·5 — 100·5

100·5 — 110·5

110·5 — 120·5

120·5 — 130·5

After fixing the class intervals and class marks for each of the classes we should proceed to consider the data. In this we should consider the ungrouped data itself as they are, without resorting to arrangement of them either in the ascending order or in the descending order. Let us consider the data in the order in which they occur and give a tally mark for the data against the appropriate class in the table.

The value 67 happens to be the first value. It lies between 60.5 to 70·5 and hence give tally mark(/) against that class. The

next value is 75 which lies between 70·5—80·5 and hence provide a tally mark for the data against the appropriate class in the table. In this manner prepare a tally mark table for all the values as explained earlier.

After exhausting all values by marking tally marks for each value, we should count the number of tally marks against the class and write the number against that class which would indicate the frequency for that class. The final position emerged would be as follows:

Class	Tally marks	Total No. of tally marks/ frequency
60·5 – 70·5	/	1
70·5 – 80·5	~~////~~	5
80·5 – 90·5	~~////~~ ////	9
90·5 – 100·5	~~////~~ ~~////~~ ////	14
100·5 – 110·5	~~////~~ ~~////~~ ~~////~~	15
110·5 – 120·5	////	4
120·5 – 130·5	//	2
		50

The total number of tally marks, ie. the total of the frequencies of all the classes will be equal to the total number of items or total number of values.

Open classes

Sometimes a few extreme values either too low or too high may occur. In such cases, open classes may be provided in the table either at the beginning or at the end to accommodate these extreme values.

Case I (Open class at the beginning)	Case II (Open class at the end)
Below 50	
50 — 60	50 — 60
60 — 70	60 — 70
70 — 80	70 — 80
80 — 90	80 — 90
90 — 100	90 — 100
100 — 110	100 — 110
110 — 120	110 — 120
120 — 130	above 120.

In order to accommodate one extreme value, say 12, we can provide an open class as 'below 50', in the beginning.

In the case of example II, we can provide an open class at the end as 'above 120' to provide for an extreme value say 320.

Cumulative Frequency Distribution (CF)

So far we have considered the frequencies of various classes. When we refer to the frequency of a particular class, we are not at all considering the frequency of other classes. But there is another frequency called 'Cumulative Frequency' which takes

into account not only the frequency of a particular class but also the frequencies of all the classes either above or below a particular class for which cumulative frequency is computed.

For example, the Cumulative frequency of a particular class say 90·5 — 100·5 is either equal to 1+5+9+14 = 29 or 2+4+15+14 = 35. There are two kinds of Cumulative frequencies, one is called 'Less than cumulative frequency' and the other is called 'Greater than cumulative frequency'.

Computation of 'less than cumulative frequency'

The less than cumulative frequency of the first class (60·5—70·5) means the number of units or the number of individuals or the number of items having the values **less than the upper limit** (70·5) of the first class, which is the same as the frequency of the first class (1).

The less than cumulative frequency of the second class (70·5—80·5) is the total number of units or items or individuals having values less than the upper limit (80.5) of the second class which means the sum total of frequency of this class and the previous class which is equal to 6 (1+ 5).

The less than cumulative frequency of the third class (80·5—90·5) is equal to the total number of units or items or individuals having values less than the upper limit (90·5) of the third class, which means the sum total of frequencies of the first three classes 15 (1 + 5 + 9) or the sum of the cumulative frequency of the previous class (6) and the frequency of the class under consideration (9) (6 + 9 = 15).

In this manner we can calculate the less than cumulative frequencies for all classes. The cumulative frequency of the last class will be equal to the total number of units or items or individuals or values in the distribution.

One important point to be noted in the calculation of the less than cumulative frequency is that it is always calculated with **reference to the upper limit of the** class.

Computation of 'greater than cumulative frequency'

The computation of greater (more) than cumulative frequency is the same as the previous one. But the main difference is that it is calculated from the bottom of the class and also with reference to the **lower limit** of the class.

The greater than cumulative frequency of the last class, 120.5 — 130.5 is equal to the number of units, or items or values or individuals whose values are **greater than** the lower limit of the class (120.5). Naturally, the greater than cumulative freqency of the last class is the same as the frequency of last class itself (2).

The greater than cumulative frequency of the next class (110.5 — 120.5) is equal to the total number of values or items or units or individuals having values greater than the lower limit (110.5) of the class. This is equal to the sum of the frequencies of this class but also all the previous classes commencing from the bottom 6 (2 + 4).

The greater than cumulative frequency of the next class (100.5 — 110.5) is equal to the total number of all the units or items or values or individuals having values greater than the lower limit (100.5) of the class. This is equal to the total of all the frequencies of the class and the previous classes at the bottom i.e., 21 (15 + 6).

In this manner we can calculate the greater than cumulative frequency of all the classes. The greater than cumulative frequency of the classes at the top will be equal to the total number of units or indivuduals or items or values of the distribution.

The *important point to be noted in the computation* of the greater than cumulative frequency is that it is always calculated with reference to the **lower limit** of the class.

The two kinds of cumulative frequencies calculated for our previous example are given in the following table.

Class	Frequency	Less than cumulative frequency	Greater than cumulative frequency	Total of the cumulative frequencies
(1)	(2)	(3)	(4)	(5)
60·5 — 70·5	1	1	1+49 = 50	1+50 = 51
70·5 —80·5	5	1+5 = 6	5+44 = 49	6+49 = 55
80·5 — 90·5	9	6+9 = 15	9+35 = 44	15+44 = 59
90·5 — 100·5	14	15+14 = 29	14+21 = 35	29+35 = 64
100·5 — 110·5	15	29+15 = 44	15+6 = 21	44+21 = 65
110·5 — 120·5	4	44+4 = 48	4+2 = 6	48+6 = 54
120·5 — 130·5	2	48+2 = 50	2	50+ 2 = 52

Sum of the less than and the greater than cumulative frequencies

We can calculate the total of these two kinds of cumulative frequencies for all classes. An interesting fact can be noticed in this. The sum of the two frequencies of any class will be equal to the total of all the units in the distribution (50) and the frequency of the respective class. This is given in the following table.

Class	Frequency	Less than c.f.	Greater than c.f.	Total of two frequencies	Total of the frequency of the class & the total frequencies of the distribution.
(1)	(2)	(3)	(4)	(5)	(6)
60·5 — 70·5	1	1	50	51	50+1
70·5 — 80·5	5	6	49	55	50+5
80·5 — 90·5	9	15	44	59	50+9
90·5 — 100·5	14	29	35	64	50+14
100·5 — 110·5	15	44	21	65	50+15
110·5 — 120·5	4	48	6	54	50+4
120·5 — 130·5	2	50	2	52	50+2

Though we may be tempted to expect the total of the two cumulative frequencies to be equal to the total number of units or values or items or individuals in the distribution, it is not actually so. This may be equal to the total, provided the two kinds of cumulative frequencies are computed with reference to one and the same particular point or limit. But this is not so here. Though we are calculating the less than cumulative frequency and greater than cumulative frequency of a particular class, both of them are not calculated with reference to one and the same limit. While the less than cumulative frequency is calculated with reference to the upper limit, the greater than cumulative frequency is calculated with reference to the lower limit of the same class.

In this process, the frequency of the particular class will be included in both the cumulative frequencies. Because of the repetition or duplication or double counting of the frequency of that class, the sum total of these two kinds of cumulative frequency exhibits such a peculiar phenomenon. This fact may help us to calculate with ease any one of the cumulative frequencies from the other, without actual computation.

Ogives

As we can draw curves for frequency distribution, we can also draw graphs for cumulative frequency and such curve is called 'Ogive'. This is only a line curve. There are two types of ogive curves, one for less than cumulative frequency and the other for greater than cumulative frequency.

The curve for the less than cumulative frequency starts at the bottom of the left and goes upwards towards right. The curve for the greater than cumulative frequency starts at the top on the left and bends downwards towards the right. Both the curves may meet at a point. A perpendicular from this point of intersection can be drawn to the X-axis. The point of intersection of the perpendicular line on the X-axis will indirectly give the median value of the distribution. The median value divides the distribution into two equal parts and that one half of them will have values less than the median value and the other half will have values greater than the median value. This we can see in greater detail when we discuss about the median in the succeeding chapters.

1. COMPUTATION OF SIMPLE FREQUENCY

From less than cumulative frequency

Class (1)	Less than c.f. (2)	Simple frequency (3)
60·5 — 70·5	1	1 — 0 = 1
70·5 — 80·5	6	6 — 1 = 5
80·5 — 90·5	15	15 — 6 = 9
90·5 — 100·5	29	29 — 15 = 14
100·5 — 110·5	44	44 — 29 = 15
110·5 — 120·5	48	48 — 44 = 4
120·5 — 130·5	50	50 — 48 = 2

As we have calculated cumulative frequency from the simple frequency, we can also calculate the simple frequency from the cumulative frequency. In the above table, the computation of simple frequency from the less than cumulative frequency is given. In column (2) of the table the less than cumulative frequencies are given. As regards the first class, we can subtract '0' from the cumulative frequency and arrive at the simple freqency. In the case of first class, both cumulative frequency and simple frequency are one and the same.

The simple frequency of any class will be equal to the difference between the less than cumulative frequency of the corresponding class and that of the previous class.

2. From greater than cumulative frequency

Class (1)			Greater than cumulative **frequency** (2)	**Simple frequency** (3)
60·5	—	70·5	50	50 — 49 = 1
70·5	—	80·5	49	49 — 44 = 5
80·5	—	90·5	44	44 — 35 = 9
90·5	—	100·5	3[illegible]	35 — 21 = 14
100·5	—	110·5	2[illegible]	21 — 6 = 15
110·5	—	120·5	[illegible]	6 — 2 = 4
120·5	—	130·5	2	2 — 0 = 2

The simple frequency can be calculated from the greater than cumulative frequency also. The simple frequency of a particular class will be equal to the difference between the greater than cumulative frequency of the class and that of the next succeeding class as given in the table above.

For the computation of either simple frequency or cumulative frequency, we require not only the classes but their respective class limits namely lower limit and upper limit of each class. Before the computation of one from the other, we should ensure that both the class limits of all the **classes should be equally defined and determined.**

Worked example

The following data give the average yield (in kg) obtained from the conduct of crop cutting experiments in 25 plots.

13, 18, 21, 21, 23, 21, 16, 16, 28, 24, 23, 26, 26, 17,

23, 24, 16, 18, 22, 22, 19, 22, 21, 12, 29.

Let us construct a frequency distribution and find cumulative frequencies for the above data.

Class interval	Tally marks	Number of tally marks
10 – 15	//	2
15 – 20	~~////~~ //	7
20 – 25	~~////~~ ~~////~~ //	12
25 – 30	////	4
		25

Class interval	frequency	Cumulative frequency	
		less than	greater than
(1)	(2)	(3)	(4)
10 – 15	2	2	25
15 – 20	7	9	23
20 – 25	12	21	16
25 – 30	4	25	4
	25		

Exercise

1. The following marks were awarded to 40 students in an examination. Construct a frequency distribution.

   30, 8, 25, 27, 48, 7, 90, 85, 28, 84, 60, 22, 25,
   33, 20, 8, 14, 17, 91, 65, 39, 40, 8, 17, 19, 90,
   36, 65, 3, 4, 75, 63, 20, 60, 80, 9, 11, 19, 6,
   70.

Draw less than and more than ogive curves for the above data.

CHAPTER VI

VITAL STATISTICS

At present, statistics on various aspects such as Agriculture, Livestock, Trade, Prices, Income are being collected. Similarly, statistical data are also collected about the people. These information on the people refer not only to the numbers but also refer to various other aspects, namely, sex, age, literacy, marital status, economic status etc. These details are collected in our country once in 10 years through the Decennial Population Census. Apart from this, details on births, deaths and diseases are collected by the State Governments, Local bodies etc. in a systematic manner as and when these incidents occur. These information are required by Government not only to decide the rate of increase in the population due to births, but also to take steps as is being done through Family Planning Welfare measures, to reduce the rate of growth of population. Similarly, information on deaths are also vitally important so that Government can take immediate measures to provide medical aids to the people whenever there is any outbreak of infectious diseases etc. Therefore, the details are collected as and when the incidents occur instead of collecting them in a periodical way, say once in a year or once in 10 years. Because of the vital importance of these statistics, they are known as Vital Statistics.

Vital Statistics refer to data on human mortality, morbidity and fertility. They provide quantitative indices of the community health, besides, they help to study the problem of growth of population.

Use of Vital Statistics

They are required by individuals for the issue of Birth and Death certificates for claims from Life Insurance Companies,

for inheritance of property etc. Marriage Certificates are required for claiming maintenance allowance and divorce certificates are required for re-marriage etc. Government may also require these data for formulating policies,to study on Manpower etc. Business people, especially, Life Insurance companies require these data to arrive at the expectation of life so as to fix the premium rates etc.

Collection of Vital Statistics

Registration of births, deaths and marriages.

The Births, Deaths, and Marriages Registration Act of 1886 provides only for the voluntary registration of vital events in India. However, a few states had passed their own acts for the compulsory registration of these events. However, there is no cumpulsory registration of marriage.

In our State, these details are recorded by the Village Officers. In the case of Municipal areas, the health authorities are recording these information. These information are sent to the District Health authorities and from these district officers the details are sent to the Director at the State headquarters. However, the system of reporting, registration and collection of Vital Statistics differ from State to State. The advantages with this type of statistics are that they are collected as and when the events occur and to that extent they are more accurate and reliable. Generally, the following data are collected regarding births and deaths.

Birth Register

1. Date of birth.
2. Name of the village.
3. Live or still birth.
4. Sex of the child.
5. Religion.
6. Age of the mother.

7. Name of the child.
8. Address of the father and mother.
9. Name and address of the Guardian.

Death Register

1. Date of Death.
2. Name of the village.
3. Name and religion of the deceased and name of the father or husband.
4. The place of residence.
5. Age, sex and occupation of the deceased.
6. Causes of death.
7. Date of last vaccination.

The statistics of births and deaths are annually collected by the Health Directorate of the Central Government. The annual reports of the State Health Directorates as well as those of major City Corporations contain more detailed information on vital events. In order to facilitate international comparison of mortality rates by specific cause, the World Health Organisation (WHO) has published the Manual of International Statistical classification of Diseases, Injuries and causes of Death. For this purpose, medical certificates are required, and they should be made compulsory. As it is not possible to adopt strictly the international scheme of classification, the Central Directorate of Health Services in our country has adopted a much simpler scheme of classification. This is given below:

Directorate General of Health Services (India) (DGHS) Classification of deaths by causes

1. Cholera 2. Small pox 3. Plague 4 (a). Fever including Malaria 4 (b). Fever-only Malaria 5. Dysentery and Diarrhoea 6 (a). Respiratory diseases including Pulmonary T.B. **6 (b). Pulmonary T.B. 7. Other causes.**

Morbidity Statistics

Morbidity relates to diseases. Some of the State Governments publish the number of patients treated in hospitals. These statistics help us to know the work-load of the hospitals. But they will not help us to get morbidity statistics.

Ad hoc surveys are undertaken by Government and private institutions. These surveys relate to morbidity, mortality, medical care. The data collected through these surveys are useful. But still they have certain limitations.

Vital Statistics Rates

Let us consider some of the Vital Statistics rates.

1. Annual Crude birth rate of a State

The annual crude birth rate of a state in a particular year is calculated by first counting the total number of births taken place in the State during the year. The total number of births is then divided by the population of the State during the year.

The population of a state is not a constant figure. The population is increasing not only day by day but also moment by moment, due to more births and less deaths. It is an ever changing phenomenon.If we take the population either at the beginning or at the end of the, year it will not give a correct picture. While the first estimate (population at the beginning of the year) is an underestimate, the second estimate (the population at the end of the year) will be an over estimation. Hence we have to strike an average of these two figures. This can be called the **Mid-year population** of the year. The total number of births in a year represents all the births taken place right from the beginning of the year upto the last moment of the year. So in all the calculation we prefer only the mid-year population to any other estimate.

$$\text{Crude Birth Rate} = \frac{\text{Total live births during the year}}{\text{Total Mid-year population of the year}}$$

The total number of births in a year will be too small, when compared with the total mid-year population. Hence the above ratio will be too small and hence, has to be expressed only in decimal places. In order to avoid decimal places and also facilitate easy remembrance this ratio is multiplied by 1000 and the result will be expressed in some convenient numbers per thousand population. Finally, a formula would emerge as follows:

$$\text{Annual Crude birth rate} = \frac{\text{Total No. of births in a year}}{\text{Total mid-year population of the year.}} \times 1000$$

i.e. $b = \frac{B}{P} \times K$

where B = Birth; P = Mid-year population; K = constant number (1000).

Calculate the Crude Birth Rate

District	Mid-year Population **in 1975**	Total birth **in 1975**
Madras	27,88,000	95,777
Chengalpattu	32,13,000	64,712
Kanyakumari	13,20,000	37,567

Crude birth rate for

(1) Madras $= \frac{95,777}{27,88,000} \times 1000 = 34.35$ per thousand.

(2) Chengalpattu $= \frac{64,712}{32,13,000} \times 1000 = 20.14$ per thousand.

(3) Kanyakumari $= \frac{37,567}{13,20,000} \times 1000 = 28.46$ per thousand.

2 Annual Crude death rate in a year

This is calculated by dividing the total number of deaths taken place in a year due to all causes by the mid-year estimated population of the year. This ratio is then multiplied by 1000 as in the case of Crude Birth rate.

$$\text{Annual Crude Death Rate} = \frac{\text{Total No. of deaths in the year due to all causes}}{\text{Total Mid Year Population of the year.}} \times 1000$$

Calculate the annual crude death rates in the following cases:

District	Mid-year population	Deaths
Madras	27,88,000	36,482
Chengalpattu	32,13,000	24,990
Kanniyakumari	13,20,000	11,710

$$\text{Annual Crude Death Rate} = \text{`}m\text{'} = \frac{D}{P} \times K$$

Where 'm' = Death rate; D = Number of deaths, P = Mid-year population; K = Constant number 1000.

Annual Crude death rate for:

1. Madras $= \frac{36,482}{27,88,000} \times 1000 = 13.09$ per thousand

2. Chengalpattu $= \frac{24,990}{32,13,000} \times 1000 = 7.78$ per thousand.

3. Kanniyakumari $= \frac{11,710}{13,20,000} \times 1000 = 8.87$ per thousand

3. Specific Mortality rate by causes of death

Mortality refers to death. Deaths are taking place due to many causes like Cholera, Small Pox, Fever etc. We can also find out the death rate due to specific cause in a year.

For this purpose the total number of deaths due to the specific cause under reference during the year will be divided by the total mid-year population. But the result will be very small to express them in decimal places. Hence this ratio will be multiplied by 100,000. Thus it will be expressed in some convenient number for lakhs of population.

$$\text{Mortality rate by cause of death} = \frac{\text{Total deaths during the year due to specific cause}}{\text{Total Mid-year population}} \times 100{,}000$$

ie: $m' = \frac{D'}{P} \times K$

where m' = death rate due to specific cause.

D' = total deaths due to specific cause.

K = Constant number (100,000).

Calculate the specific death rates in the following cases for Tamil Nadu in 1975.

Causes	Total No. of deaths
1. Fever	64,787
2. Dysentery etc.	20,997
3. Respiratory diseases	32,315

Total mid-year population of Tamil Nadu in 1975 = **44428,000.**

(i) Specific death rate due to fever $= \frac{64787}{44428{,}000} \times 100\,000 = 145.82$

$=$ 145.82 per lakh of population.

(ii) Specific death rate due to Dysentery $= \frac{20,997}{44428,000} \times 100,000$

$= 47.26$ per lakh of population.

(iii) Specific mortality rate due to respiratory diseases $= \frac{32,315}{44428,000} \times 100,000$

$= 72.73$ per lakh of population.

4. Specific Mortality rate by age and sex

$$= \frac{\text{Total deaths during the year among persons of specified age and sex}}{\text{Total mid-year population of that specified age and sex.}} \times 1000$$

Suppose we want to find the death rate of population in the age group 40-59 in a year, this should be calculated as follows:

$$\frac{\text{Total no. of deaths in the year in the age group 40-59}}{\text{Total mid-year population of the people in the age group 40-59 in the year.}} \times 1000$$

Mid-year population in the age group 40-59	Total number of deaths in the year in the age group 40-59
149,000	3057

Death rate in the age group 40-59 $= \frac{3057}{149,000} \times 1000 = 20.52$ per thousand.

This can be further split up into two parts. The population consists of males and females. Therefore, we can calculate the death rate among males in the age group 40-59. The method is same. The only difference is that we should divide the total number of deaths among males in the age group 40-59 by the total mid-year population of Males in the age group 40-59.

	Mid-year population	No. of deaths
Males	79,000	1,735
Females	70,000	1,322
	149,000	3,057

(i) Death rate among males in the age group 40-59 $= \dfrac{\text{No. of deaths among males in the age group 40-59}}{\text{Total mid-year population of males in the age group}} \times 1000$

$= \dfrac{1,735}{79000} \times 1,000$

$= 21.96$ per thousand.

(ii) Death rate among females in the age group 40—59 $= \dfrac{\text{No. of deaths among females in the age group 40-59}}{\text{Mid-year population in the age group}} \times 1000$

$= \dfrac{1,322}{70,000} \times 1,000$

$= 18{\cdot}9.$ per thousand.

5. Infant mortality rates

$$\frac{\text{Total No. of deaths during the year among infants under one year age}}{\text{Total number of live births in the year.}} \times 1,000$$

ie: $\dfrac{D_o}{B_o} \times K$

Where D_o = Death below 1 year, B_o = Total live births.

K = Constant number 1,000.

Total number of live births in Madras in 1975 = 95,577

Total number of infant deaths in Madras in 1975 = 8,259

$$\text{Therefore, infant mortality rate} = \frac{8{,}259}{95{,}577} \times 1000$$

$$= 86.41 \text{ per thousand.}$$

The formula is self explanatory.

6. Still birth rate

Still birth relates to birth of children born dead.

$$\text{Still Birth Rate} = \frac{\text{Total still births in the year}}{\text{Total births (Live births + Still Births) in the year.}} \times 1{,}000$$

$$\text{i.e. : } \frac{S}{S+B} \times K$$

Where S = Still Birth, B = Live birth

K = Constant (1,000).

7. Maternal Mortality Rate

Maternal mortality rate relates to death rate due to delivery of child.

$$= \frac{\text{Total No. of deaths due to puerperal causes in the year}}{\text{Total births (Live births+Still births) during the year.}} \times 1{,}000$$

$$= \frac{D^P}{S+B} \times K$$

Where D^P = Death due to puerperal causes

S = Still birth; B = Live birth; K = Constant (1,000).

Exercise

1. Explain the uses of vital statistics and explain the method of collection of vital statistics.

2. What are the different vital statistics rates?

3. Write short notes on:

   i. Annual crude birth rate and crude death rate.

   ii. Specific mortality rate.

   iii. Maternal mortality rate.

CHAPTER VII

MORTALITY TABLES OR LIFE TABLES

Life Insurance has become more familiar nowadays not only among the urban population but also among the rural population. Under this system, the company agrees to pay a specifed amount to the individual on the completion of a particular age or on the death of the individual, whichever is earlier. In turn, the individual has to pay a certain specified amount at periodic intervals, which is known as premium. These premiums are calculated on the basis of the chance (Probability) that a person of a particular age, say X, will live for some more years, say 'n' years. By undertaking such risk among a large number of population, the insurance companies are able to earn a profit after meeting all the payments on account of risk and other establishment charges etc. The table that enables us to find out the chance (probability) that a person at a particular age will survive for a specific number of years is known as Mortality Table or Life Table.

Life Table

Life table is a document which reflects the course of mortality of a population. A life table comprises of a number of columns. Let us consider the table given below:

x (1)	l_x (2)	d_x (3)	q_x (4)	e_x (5)
0 l_0	100,000	15,000	150.0	
1 l_1	85,000	4,000	47.1	
2 l_2	81,000	2,000	24.7	
3 l_3	79,000	1,000	12.7	
4 l_4	78,000	5,00	6.4	
5 l_5	77,500	3,00	3.9	

Age - x

The first column is the column specifying age (x), namely 0, 1, 2, 3..........n.. In our example, the total number of persons at the age 0 is equal to 100,000. The number of persons at age 1 is equal to 85,000. Similarly the number of persons who have completed 2 years or attained 2 years is 81,000.

Number of persons living (l_x)

The second column is the survival column, which may be considered as the basic column of the life table. It gives the number of persons surviving to successive ages of life out of a group of persons born simultaneously. Usually a convenient number say 100,000 of persons are chosen to comprise this group. This number is called the Radix of the table.

At successive birth day, the number of persons in the group is reduced by death. The number of persons surviving at an age x is denoted by the symbol l_x, which is an abridged form of 'living at 'x'. The principle involved is very simple. It traces the survivalship history of a group of persons from birth until all the persons in the group are accounted by death.

Number of deaths -dx

The third column of the life table gives the number of deaths taken place between two consecutive years. It is clear from column (2) of the life table, that the difference between two consecutive numbers in the survival column will give the number of deaths taken place between two consecutive birth days (ages) of persons in the group. The symbol employed to represent the number of persons dying between the ages x and $(x+1)$ is 'dx'.

In our example, the number of persons at '0' age is 100,000 The number of persons who have attained one year of age is 85,00((vide col. 2). It means 15,000 (100,000—85,000) persons hav died within the one year period. If l_0 denotes 100,000 and '$d_($ denotes 15,000, we can construct the following equation:

$$d_0 = l_0 - l_1$$

Similarly, $d_1 = l_1 - l_2$

$$d_2 = l_2 - l_3$$

The general formula is $d_x = l_x - l_{(x+1)}$

Since all the persons who attain the age 'x' will ultimately die, the sum of all the entries in the 'dx' column from the age x to the end of the table will be equal to the number of persons living at the age x which is equal to l_x.

ie: $l_x = d_x + d_{x+1} + d_{x+2} + \dots\dots + d_{x+n}$

$$= \sum_{x=1}^{n} d_x$$

Rate of mortality (q_x)

The fourth column in the life table is the column of mortality rate. This column gives 'q_x' which may be defined as the chance (probability) of a person aged x dying within a year. Usually mortality rates are expressed in rates per 1,000 ($q_x \times 1{,}000$) in order to avoid decimals.

In our example, the number of persons at age 0 is 100,000. The number of persons who have died before attaining the first year is 15,000. Therefore, the chance of a person at the age '0' to die before attaining the age of 1 is equal to 15,000/100000 = 0·15. This is denoted by the letter q_x. In order to avoid the decimal place, it is multiplied by 1,000 and expressed in integers as 150 per thousand.

$$\text{ie: } q_x = \frac{l_x - l_{x+1}}{l_x} \times 1{,}000 = \frac{d_x}{l_x} \times 1{,}000$$

The complementary chance (probability) namely 'p_x' is equal to 1 - q_x gives the chance (probability) of a person aged x surviving one year more and is equal to $\frac{l_{x+1}}{l_x}$

In our example, the number of persons at age '0' is 100,000. The number of persons who have attained one year or lived upto one year is 85,000. Therefore, the chance of a person at '0' age to complete one year is euqal to

$$\frac{85000}{100000} = \frac{l_{x+1}}{l_x} = 0.85$$

We know that the chance of a person at '0' to die before attaining one year is 0·15. Naturally, the chance of a person at '0' to complete 1 year is just complementary of (1 – 0·15) = 0·85. This is due to the fact that the total of chances of a person at '0' age either to attain one year or to die before attaining one year is equal to 1.

ie: If P_x denotes the chance of a person living and q_x denotes the chance of a person dying within a particular period (ie: one year), the following equation will emerge.

$p_x + q_x = 1.$

Therefore, $p_x = 1 - q_x$

Therefore, $q_x = 1 - p_x$

Hence p_x and q_x are said to be complementary to each other.

Expected Life (e^x)

The fifth column of the Life Table gives the expected life of a person aged x. It shows at each age, the average number of years will still be remaining to a person of that age until death. The expectation of life may be defined as a ratio between the total number of years lived by all the persons who have attained an age x, divided by the total number of persons who are actually at x.

The expectation of life can be derived by considering the number of completed years lived by the persons surviving to an age $(x+1)$. $l_{(x+1)}$ persons live throughout the one year from x to $(x+1)$ and hence the number of years lived by these $l_{(x+1)}$ persons in the first year is $l_{(x+1)}$ years. But d_x persons die

during the year. Assuming that death takes place at a uniform rate throughout the year, the number of years lived by these dx persons will be $dx/2$ years or $\frac{l_x - l_{(x+1)}}{2}$. Therefore, the total number of years lived by the persons aged x in the first year (from x to $(x+1)$ year) is $l_{x+1} + \frac{dx}{2}$

$$= l_{(x+1)} + \frac{l_x - l_{(x+1)}}{2}$$

$$= \frac{2\, l_{(x+1)} + l_x - l_{(x+1)}}{2}$$

$$= \frac{l_x + l_{(x+1)}}{2} \text{ years}$$

In this manner, we calculate the number of years lived by the persons aged x upto $l_{(x+8)}$ years. This will be

$$= \frac{l_x + l_{(x+1)}}{2} + \frac{l_{(x+1)} + l_{(x+2)}}{2} + \frac{l_{(x+2)} + l_{(x+8)}}{2} + \ldots\ldots$$

$$= \frac{l_{x+2} l_{x+1}\ l_{(x+1)} + 2\, l_{(x+2)} + 2\, l_{(x+8)}}{2} + \ldots\ldots$$

$$= \frac{l_x}{2} + l_{(x+1)} + l_{(x+2)} + l_{(x+8)} \text{ years} + \ldots.$$

Let us explain this with the help of figures in our table.

Out of the 100,000 persons at '0' age, only 85,000 have attained one year.

Therefore, the number of years lived by the

85000 persons from 0 to 1 year $= 85000 \times 1 = 85000$

$$= l_{(x+1)} \times 1 = l_{(x+1)}$$

Out of 100,000 persons at '0' age, 15,000 persons have died before attaining one year. These 15,000 persons have not died all of a sudden at a point of time. Instead, the death is taking place every moment right from '0' to 1 year. Some of them would have lived upto one month, some others upto 11 months. It can be assumed that these 15,000 (dx) persons who have died within one year have lived for $\frac{1}{2}$ year each. Therefore, the number of years lived by these 15,000 persons who have died within one year (0 to 1 is

$$\frac{15000}{2} = 7,500 \text{ years.}$$

$$\text{(or) } \frac{15000}{2} = \frac{100,000 - 85,000}{2}$$

$$\text{ie: } \frac{dx}{2} = \frac{l_x - l_{(x+1)}}{2}$$

the total number of years lived by $l_{(1+x)}$ (85000) persons who have lived for one year each = 85000 years.

$$= l_{(x+1)} \times 1 \text{ years.}$$

$$= l_{(x+1)} \text{ years.}$$

dx persons have lived for $\frac{l_x - l_{(x+1)}}{2}$ years.

$$= \frac{100,000 - 85,000}{2}$$

Therefore, the total number of years lived by 100,000 persons at '0' age at the time of attaining one year is

$$85,000 + 7,500 = 92,500 \text{ years.}$$

$$\text{or } l_{(x+1)} + \frac{l_x - l_{(x+1)}}{2}$$

$$= \frac{2\, l_{(x+1)} + l_x - l_{(x+1)}}{2}$$

$$= \frac{l_x + l_{(x+1)}}{2}$$

$$= \frac{100,000 + 85,000}{2} = \frac{185,000}{2} = 92,500 \text{ years}$$

In the same manner we can calculate for the remaining years as:

$$\frac{l_x + l_{(x+1)}}{2} + \frac{l_{(x+1)} + l_{(x+2)}}{2} + \frac{l_{(x+2)} + l_{(x+3)}}{2} + \frac{l_{(x+3)} + l_{(x+4)}}{2} + \frac{l_{(x+4)} + l_{(x+5)}}{2}$$

$$= \frac{100{,}000 + 85000}{2} + \frac{85000 + 81000}{2} + \frac{81000 + 79000}{2} + \frac{79000 + 78000}{2} + \frac{78000 + 77500}{2}$$

$$= \frac{100{,}000}{2} + \frac{2 \times 85000}{2} + \frac{2 \times 81000}{2} + \frac{2 \times 79000}{2} + \frac{2 \times 78000}{2} + \frac{2 \times 77500}{2}$$

$$= \frac{100{,}000}{2} + 85000 + 81000 + 79000 + 78000 + 77500.$$

$$= \frac{l_x}{2} + l_{(x+1)} + l_{(x+2)} + l_{(x+3)} + l_{(x+4)} + l_{(x+5)}.$$

The average number of years lived by persons aged x is arrived at by dividing the total number of years lived by the number of persons at age x.

$$\frac{\frac{l_x}{2} + l_{(x+1)} + l_{(x+2)} + l_{(x+3)}}{l_x}$$

$$= \tfrac{1}{2} + \frac{l_{(x+1)} + l_{(x+2)} + l_{(x+3)}}{l_x}$$

This can be written as follows:

The total number of years lived by the number of persons at '0' age till the attainment of 5 yeras of age is:

$$= \frac{100{,}000}{2} + 85000 + 81000 + 79000 + 78000 + 77500.$$

The total number of persons at age 0 = 100,000,

Therefore, the average number of years lived by each person at x '0' age

$$= \frac{\frac{100,000}{2} + 85,000 + 81000 + 79000 + 78000 + 77500}{100,000}$$

$$e^x = \frac{\frac{l_x}{2} + l_{(x+1)} + l_{(x+2)} + l_{(x+3)} + l_{(x+4)} + l_{(x+5)}}{l_x}$$

$$e^x = \frac{1}{2} + \frac{l_{(x+1)} + l_{(x+2)} + l_{(x+3)} + l_{(x+4)} + l_{(x+5)}}{l_x}$$

This is nothing but equal to

$$e^x = \frac{\frac{l_x + l_{(x+1)}}{2} + \frac{l_{(x+1)} + l_{(x+2)}}{2} + \frac{l_{(x+2)} + l_{(x+3)}}{2} + \frac{l_{(x+3)} + l_{(x+4)}}{2} + \frac{l_{(x+4)} + l_{(x+5)}}{2}}{l_x}$$

This is called the expectation of life and is denoted by 'e^x'.

We can rewrite the formula as:

The symbol L_x is used to denote the number of years lived by l_x persons between the age x and $(x+1)$. where

$$Lx = \frac{l_x + l_{x+1}}{2} \text{ and } L_{(x+1)} = \frac{l_{(x+1)} + l_{(x+2)}}{2} \text{ etc.}$$

$$Tx = L_x + L_{(x+1)} + L_{(x+2)} \text{ etc.}$$

$$e^x = \frac{L_x + L_{(x+1)} + L_{(x+2)}}{lx} = \frac{T_x}{l_x}$$

Construction of Life Table

The fundamental data necessary for the construction of life table based on population data are:

(1) an enumeration of the population at a given time classified according to age.

(2) The number of deaths within the population in a period of one year and also classified according to age.

The function q_x ie: col (4) is determined indirectly through the average death rate 'm_x' among persons of ages between x and $x+1$.

d_x denotes the number of annual deaths between these ages. Then $m_x = \frac{dx}{Px}$ will denote the average rate of mortality. i.e. the central death rate per head of the population (P_x) comprised within the age group. We can find out the q_x from the observed quantity m_x.

We know that $q_x = \frac{d_x}{l_x}$

The average number of persons living during the years between the ages x and $x+1$ will be,

$$\frac{l_x + l_{x+1}}{2} \text{ or } lx - \frac{d_x}{2} \text{ or } l_{x+1} + \frac{d_x}{2}$$

Let us assume that the death takes place at an uniform rate during the year. Therefore, the average death rate m_x will be

$$m_x = \frac{dx}{lx - dx/2} = \frac{dx}{l_{x+1} + \frac{dx}{2}}$$

But $q_x = \frac{d_x}{l_x}$

Therefore $l_x \times q_x = d_x$.

Substituting the value of $l_x \times q_x$ for d_x in the above equation for m_x we get the following three formulae:

1. $m_x = \frac{2\,q_x}{2 - q_x}$

2. $q_x = \dfrac{2\, m_x}{2 + m_x}$

3. $p_x = 1 - q_x$

$$= \frac{2 - m_x}{2 + m_x}$$

Exercise

1. Explain the various columns of mortality table.
2. Calculate the probability of a man aged 40 dying at the age of 45 from the following table.

Age	Numbers of persons living
40	10000
41	9500
42	8800
43	8000
44	7500
45	7200
46	6500
47	5000
48	3500
49	2700
50	2000

3. Write short notes on:
   i. Expectation of life.
   ii. Mortality rate.
4. **Construct a mortality table.**

CHAPTER VIII

STATISTICAL SYSTEM IN INDIA

The nature and structure of the statistical organisation in India is governed by the constitutional set up under the Indian Constitution. This reponsibility is shared by threefold classifications of subjects, Foreign Trade, Banking and Currency and Population which are wholly allocated to the Central Goverment. Industry comes under the common category allotted to both the Centre and the State. Agriculture, Education are State subjects. Even in the field of Agriculture the centre is acting as a Co-ordinating agency. At the centre and State, each Ministry and each department has a statistical wing to collect and process the data relating to their activities besides the central agency for the collection and presentation of data for the Centre and State, as a whole. A detailed description of the function of the various statistical organisations and their publications is given in the following paragraphs.

Department of Statistics in the Ministry of Planning

In view of the growing importance of satistics, Government had constituted the Department of Statistics in 1951. The co-ordination of the statistical activities of the various Central and State agencies, the setting up of standards and norms in connection with the collection and presentation of data are the important functions of this organisation. The Central Statistical Organisation and the Directorate of National Sample Surveys are the main Statistical Units under this Department of Statistics. It is closely associated with the Planning Commission.

CENTRAL STATISTICAL ORGANISATION (C.S.O.)

The Central Statistical Organisation was set up in 1951. Its functions consist of (1) Co-ordination of the Statistical work of different Ministries and other Government agencies with a view to eliminate and prevent unnecessary duplication and thereby reduce overall cost. (2) Development of definition and standards for improving national and international comparability and improvement of the quality of the data. (3) Giving advices to other similar organisations in the Government and arrange inter-departmental discussions and co-operation. (4) Keeping close touch with similar wings in other countries and world organisations in the development of new methodology etc. (5) Serving as a channel of communication with the United Nations Statistical Organisation. (6) Publication of Annual Abstract of Statistics, Monthly Abstract of Statistics, Weekly supplement to Monthly Abstract of Statistics, Estimates of National Income etc.

Undertaking Statistical work relating to planning, Estimation of Annual National Income and conduct of research in National Income, Conduct of training courses in official statistics and undertaking studies concerning population and demography are some of the other functions of this Organisation. The following are important regular publications:

1. Statistical Abstract of India — Annual
2. Statistical Hand Book of India— "
3. Estimation of National Income— "
4. Sample surveys of current interest— "
5. Basic Statistics relating to— Indian Economy. "
6. Monthly Abstract of Statistics — Monthly
7. Weekly supplement to the Monthly Abstract of Statistics.

Industrial Statistical wing of the Central Statistical Organisation

Its main function is to plan, guide and co-ordinate the

1. Annual Survey of Industries under the Collection of Statistics Act

2. Collection of monthly statistics of production of selected Industries and their publication.

3. Computation of monthly index of industrial production in India, etc. The publications are

(i) Annual Survey of Industries.

(ii) Monthly Statistics of industrial production.

(iii) Index of Industrial Production.

Directorate of National Sample Surveys

It was established in 1950. Its main functions are collection of statistics required for National Planning and National Income Estimation, Conduct of sample surveys for the collection of data on Socio-Economic Conditions of the population, prices, wages, production in small scale industries, conduct of sample surveys in organised industrial sectors. They are publishing the reports of the National Sample Surveys conducted by them.

Department of Commercial Intelligence and Statistics

The unit was established in 1895. The function of this organisation is to compile data on trade- foreign, coastal, rail and river-borne and data on maritime navigation and customs and excise revenue.

The above functions at present cover the following items of activities:

1. Collection and furnishing of commercial information required by Government and the trade.

2. Maintenance of a register of firms in India and also relevant information relating to the firms.

3. Disseminate commercial information received from the Indian Government trade representations abroad.

4. Compilation and publication of statistics of trade, shipping etc. and publish the weekly Indian Trade Journal.

The various publications of this unit and their periodicity are as follows:

1. Monthly statistics of Foreign Trade of India - Imports and Exports	—	Monthly
2. Supplement to the Monthly Statistics of Foreign Trade of India.	—	Monthly and Quarterly.
3. Customs and Excise revenue statistics of India	—	Monthly
4. Statistics of Coastal Trading in India	—	Quarterly
5. Accounts relating to the Inland (Rail and River borne) Trade of India.	—	Quarterly
6. Statistics of the Maritime Navigation of India.	—	Annual
7. Annual Statement of Foreign Trade	—	Annual
8. Annual Statistics of the Foreign Trade of India by Customs.	—	Annual

Office of the Economic Adviser

Its main functions are to maintain wholesale prices and prices data in general. It also acts as the co-ordinating unit. It publishes the weekly Index Numbers of wholesale prices in India.

Department of Statistics, Reserve Bank of India

This unit is responsible for the compilation of statistical data for the Banks Bulletin and its weekly supplement and for the supply of statistics on banking, currency, finance etc. This department also publishes a series of index numbers of prices of gilt edged and industrial securities and yields at It collaborates in the surveys undertaken by the other departments of the bank. Analysis of the company accounts which is an important work is done in this Department.

The division of International Finance of the Bank's Economic Department is incharge of the compilation and refinement of India's balance of payment Statistics, data on visible and invisible trade, capital transaction, foreign exchange reserves and changes in the country's foreign assets. The Agricultural credit department of the Bank attends to compilation and publication of statistics relating to co-operative institutions.

The publications are:

1.	Statistical Tables relating to Banks of India. Report on Currency and Finance.	Annual
2.	Statistical Statements relating to co-operative Movements in India.	Annual
3.	Trend and progress of Banking in India.	Annual
4.	Reserve Bank of India Bulletin	— Monthly
5.	Weekly Statistical Supplement to Reserve Bank of India Bulletin.	— Weekly

Directorate of Economics and Statistics in the Ministry of Food and Agriculture

This Directorate is responsible for collection, compilation and publication of all agricultural statistics. The data covering fields of Agriculture, Livestock, Fisheries, Forests, are

collected from State Departments. It also collects data on agro-economic policy. It issues a series of regular publicatic ns which will be of great value to Government and Research workers.

Annual Publications

1. Estimates of area and production of principal crops in India.
2. Area and production and average yield per hectare of food crops.
3. Indian Agricultural Statistics.
4. Indian Land Revenue Statistics.
5. Indian Agriculture.
6. Indian Livestock Statistics.
7. Agricultural prices in India.
8. Agricultural wages in India.
9. Publication of Coffee, Cotton, Jute, Lace, Oilseeds, Sugar, Tea, Tobacco, Rubber in India.
10. Indian Forests Statistics.
11. Indan Livestock Statistics (Quinquennial.)
12. Indian Agricultural Atlas.
13. Bulletin of Agricultural prices — Weekly.

Institute of Agricultural Research Statistics (IARS)

This was formerly a wing of the Indian Council of Agricultural Research (ICAR). The functions of this Institute are to advise

on planning of experiments in Agriculture and Animal Husbandry, to impart training in Agriculture and Animal Husbandry Statistics, to carry out fundamental research on the application of the statistical methods to agriculture and animal husbandry problems and to conduct research in sampling techniques for collection of agriculture and animal husbandry data.

Statistical Bureau in the Directorate General of Health Services

Its main functions are to centralise collection, compilation, analysis, evaluation and dissemination of all information on health statistics for the nation as a whole, to collaborate with the Registrar General and other State departments on maintenance of health statistics, to disseminate epidemic intelligence etc. Health Statistics of India which is an annual publication is issued by this unit.

Office of the Registrar General of India

It deals with the population statistics including vital statistics and census. Vital statistics of India, an annual publication besides Indian population Bulletin on half yearly basis are the important publications. Decennial Statistical publication on population census is another important document released by this office.

Labour Bureau

The main functions are (1) Collection, compilation and publication of labour statistics, (2) Construction and maintenance of Consumer Price Index numbers, (3) Conducting research in specific problems with a view to furnishing data for the formulation of labour policies and bring out statistical publications relating to labour affairs.

Indian Labour Year Book, Indian Labour Statistics, Trade Union in India, Large Industrial Establishment, Minimum wages

under the Minimum Wages Act are the important Annual Publications. Indian Labour Journal is published as a monthly issue.

Directorate General of Employment and Training

The statistical wing of the Directorate is in-charge of collection, compilation, and analysis of data relating to the activities of Employment Exchanges, training schemes, census of Central Government Employees according to the pay range and processing of employment marketing information. Report on the Census of Central Government employees is published as an annual issue.

Indian Bureau of Mines

The functions are (1) collection, compilation and interpretation of mineral statistics, (2) collection and maintenance of information regarding world mineral production, prices and trade; (3) compilation of indices of mineral production and prices, preparation of estimates of national income originiting from the mining sector, (4) Publication of statistics relating to mineral production in India, mineral trade and indigenous consumption etc.

Mineral Year Book is an annual publication in addition to the 'Statistical Survey of Mineral Production' which is being issued on a monthly basis.

Programme Evaluation Organisation

The statistical unit of this organisation looks after the statistical aspects of planning and conduct of evaluation studies, and publication of evaluation reports and assist the Planning Commission.

Statistical Directorate in the Railway Board

Compilation of statistics of Railways can come under two categories namely the statistics required by the Railway Board

to assess the activities of the Railways and the other is the collection of the data required by each division for their own domestic purposes.

Statistical organisations of varying strength exist in different railways. The broad functions of these organisations are the collection, compilation and analysis and dissemination of statistical data relating to the various phases of working of the Railways, such as operation, commercial, accidents, traffic., etc.

Various publications giving statistical details are published on monthly, quarterly and annual basis.

Statistical section in the Ministry of Transport (Road Wing)

The main functions are collection, compilation, analysis, interpretation of statistics concerning all aspects of road transport, designing experiments to assess the cost of operation of motor vehicles on different types of roads, conducting traffic and economic surveys in rural areas for assessing requirements and studying road economics. 'Basic Road Statistics of India' is published annually besides 'Road Facts in India' which is a quinquennial publication.

Indian Meterological Department

The main functions are the scruitny and analysis of the data received from the Observatories in India and from rain-guage stations maintained by States. They publish weekly, monthly and annual weather reports, and also publish development of new methods and techniques for analysis and presentation of data for weather forecasts.

National Building Organisation

The main functions of this unit are development of methodology and organisation of statistics of housing and buidling construction, collection, co-ordination and presentation of data at the National Level.

In addition to the above statistical wings, there are statistical wings in different departments like Tea Board, Coffee Board, Coir Board, Defence departments etc. They also collect valuable data and compile them and make available for Government, Planning Commission etc.

State Statistical Bureau

Every State Government have also set up statistical departments in their States. They mainly collect statistics relating to Agriculture, Prices, Trade, Wages etc. They also conduct Crop Cutting Experiments on important crops to estimate the yield. They publish Season and Crop reports every year. Besides, they publish weekly wholesale and retail prices of selected commodities prevailing in important centres. They also compile cost of living index for important urban centres They undertake compilation of state income. They conduct periodic Socio-Economic Surveys and evaluation studies.

Besides the Statistics Departments, every department has also set up Statistical Cell in their departments to collect, compile statistical data relating to their departments for framing their policies and day to day administration.

Exercise

1. Write an essay on the statistical system in India.

TAMILNADU TEXTBOOK SOCIETY

LEARN THOROUGHLY